WORKING PAPERS

N° 1567

January 2024

"Antibiotic Stewardship in Primary Care: Evidence from Pay-for-Performance in France"

Gökçe Gökkoca

Toulouse School of Economics

Antibiotic Stewardship in Primary Care: Evidence from Pay-for-Performance in France

Gökçe Gökkoca*

January 24, 2024†

Link to the latest version

Abstract

This paper investigates whether financial incentives for curbing antibiotic prescriptions are effective and how the design of incentives plays a role in influencing physician behavior. Using prescription-level data from French general practitioners over six years, I provide evidence of the incentives' effectiveness by exploiting variation in the set of diseases that the physicians treat as well as in the reward scheme. To understand how they respond, I propose a model that incorporates financial incentives into physician's decision-making and test the predictions of the model. The results highlight that the reduction in antibiotic prescriptions varies across different diseases, in line with the physicians' altruism and, hence, the patient's needs. Moreover, forward-looking physicians are influenced by the marginal cost of antibiotic prescriptions and the design of the incentives. While the program is effective, the magnitude is moderate, with a 2 percentage point drop in the antibiotic prescription rate. Comparing the effect to the cost of the program, conditioning the rewards on prescription rates rather than the improvement over time plays a role. As a result, while aggregate bonus payments per physician remain modest (on average 0.2% of physicians' annual income), the cost per avoided prescription is substantial (on average 56% of the fixed visit fee).

Keywords: antibiotic stewardship, pay-for-performance, physician behavior

*Toulouse School of Economics, Toulouse, France, email: gokce.gokkoca@tse-fr.eu

†I thank Pierre Dubois for his support and guidance, as well as Léa Bignon, Céline Bonnet, Isis Durrmeyer, Gaston Illanes, Bruno Jullien, Mathias Reynaert, François Salanié, and Molly Schnell for helpful comments and discussions. I also thank seminar participants at the Toulouse School of Economics and Northwestern University and participants at the Workshop on the Economics of Antibiotics 2021 in Toulouse for insightful comments. The author gratefully acknowledges funding from the Agence Nationale de la Recherche under grant Antibionomics - 16-CE36-0004-02 and ANR-17-EURE-0010 (Investissements d'Avenir program), funding under the "PSPC" call for projects operated on behalf of the French government by Bpifrance as part of the "investments for the future" program (PIA) ARPEGE and funding from the TSE-P Health Center whose lists of sponsors can be found at https://www.tse-fr.eu/health. I thank GERS CEGEDIM for access to the THIN data.

1 Introduction

Antibiotics are developed to treat bacterial infections. However, they are not permanent cures as bacteria evolve to increase their defenses. Antibiotic resistance occurs when bacteria no longer respond to the antibiotic treatment (i.e., become resistant). As this phenomenon redefines and narrows the scope of treatable infections, it presents an immense threat to global health and development. Antibiotic stewardship is a key component against antimicrobial resistance since consumption is a major driving force of resistance (Fleming-Dutra et al. (2016); Ventola (2015); WHO (2015)). Moreover, inappropriate use of antibiotics is quite common. The estimates for the share of unnecessary prescriptions are approximately 30% for the US and 50% for France (Carlet and Le Coz (2015); Fleming-Dutra et al. (2016)). In many countries, antibiotics are prescription medicines and are most frequently prescribed in the outpatient setting (HSA (2022); Simon et al. (2022); Thilly et al. (2020)). Hence, shaping the prescription behavior of physicians in the outpatient setting for better use of antibiotics is important in managing consumption externalities.

In this paper, I consider the case of France and examine the effectiveness of financial incentives to general practitioners in curbing antibiotic prescriptions and the influence of the incentive design. France has been struggling with high resistance rates and consumption levels[1]. Since the early 2000s, there have been two large information campaigns and updates to guidelines in line with the evolution of resistance. The efficacy of these policies in improving physician prescription practices is challenged in Carlet et al. (2020). Looking at prescription trends between 2001 and 2018, the authors find that the use of antibiotics in human medicine bounced back after the information campaigns and failed to alter the trend. These facts highlight why antibiotics are in the pay-for-performance (P4P) program introduced in 2012[2]. The success of financial incentives for physicians concerns public health in addition to public finance. Therefore, it is crucial to understand if the financial incentives work and how they work. Pay-for-performance in the outpatient setting has been increasingly utilized to harmonize physician performance and to encourage better practices[3]. In the case of antibiotics, there is a discrepancy between the social cost of resistance and the cost internalized by the physician at the time of antibiotic prescription. The goal of the pay-for-performance program is to address this discrepancy leading to over-prescription. Hence, the French Health Authority uses contracting upon two observable actions and rewards physicians upon good performance. These actions are (i) the prescription rate of antibiotics to low-risk patients and (ii) the proportion of antibiotics that are useful for resistant bacteria.

The effectiveness of P4P programs remains an open question partly due to the selection of performance

[1] Antibiotic consumption at defined-daily-dose remains around 30% higher than the European average (DiMasi et al. (2010)).

[2] The program is called Remuneration based on public health objectives (ROSP). The performance criteria are selected with public health objectives in mind.

[3] Examples include the United Kingdom (NHS's Quality and Outcomes Framework - QOF), Australia (Practice Incentives Program - PIP), Sweden, Taiwan, and China.

measures, bonus design, and the relevant population of patients and diseases (Van Herck et al. (2010)). These features of a P4P program could affect the returns if not properly addressed. How to tailor the incentives to generate better results remains an open empirical question (Eijkenaar et al. (2013)). The lack of evidence also applies to financial strategies around antibiotic prescriptions. Yoshikawa et al. (2021) provides a summary of different ways of incentivizing physicians under antibiotic stewardship programs and highlights the need to understand the mechanisms behind provider response in prescription patterns. This paper asks if the incentives are working in a socially desired way and, more specifically, if bonus design is important for the success of the P4P program. I rely on prescription-level data from a panel of general practitioners in France from 2014 to 2019 to study these questions. For the physicians in my sample, I observe every prescription with the corresponding treatment, date and patient identifier. The P4P program defines two performance criteria concerning the use of antibiotics: (i) "Overall Antibiotics Criterion" (OAC) and (ii) "At-risk Antibiotics Criterion" (AAC). The OAC aims to decrease antibiotic prescription rates among (low-risk) patients. It targets the inappropriate use of all antibiotics and does not differentiate among the types of antibiotics. The AAC aims to decrease the share of a particular group of antibiotics (a subset of broad-spectrum antibiotics) among all antibiotic prescriptions. It is more targeted to protect the efficacy of high-value antibiotics against resistant bacteria, which have been seeing rising levels of resistance. Thus, it encourages physicians to not overly prescribe these specific antibiotics where there are other viable options.

In the first part of the paper, I show that both criteria had an impact on prescriptions. I exploit the changes in the design of the bonus function over time together with the heterogeneity in the pool of diseases the physicians treat to show that the introduction of more stringent requirements for performance pay caused physicians to increase their performance. Studying the OAC, I rely on the heterogeneous prevalence of viral infection across regions. I focus particularly on bronchitis infections, a common reason for unjustified antibiotic prescriptions. I find a more pronounced decrease in prescription rates by physicians in high viral infection-prone regions. This finding supports the hypothesis that if the incentives lead to a decrease in the antibiotic prescription rate, physicians practicing in regions with a high prevalence of viral infections, which often lead to unwarranted antibiotic prescriptions for conditions such as bronchitis or the flu, experience a greater reduction in their prescription rates compared to those working in areas with lower viral infection rates. Therefore, the incentives are more effective when avoiding antibiotic treatment does not pose a threat to patient health. In the final section of the paper, I show that physicians indeed do not cut their prescriptions where they are sure that antibiotic prescription is needed. I use a similar approach to demonstrate the effectiveness of the AAC, which was introduced in 2017. It was anticipated that the AAC would be particularly beneficial for physicians treating conditions that necessitate antibiotics. Refraining from prescribing antibiotics when they are needed is harmful to the patient. However, if there is an excessive

use of "at-risk" antibiotics, physicians can make adjustments at the intensive margin and prescribe alternative antibiotics. The findings support a more substantial performance improvement, indicated by a decrease in the share of "at-risk" antibiotics, among physicians who treat a higher number of cases requiring antibiotic treatment.

In the second part of the paper, I suggest a model of treatment choice that incorporates the P4P scheme. I allow financial incentives to enter the physician's utility through income. Moreover, physicians derive utility from the patients' utility (depends on antibiotic treatment) through an altruistic term and have private benefits attached to antibiotic prescriptions[4]). The main components of the model are in parallel to the literature on physician decision-making (Dranove (1988)). The model also employs varying health benefits to patients based on the severity of the infection in a similar way to Schnell (2017). One particular feature of modeling antibiotic prescription decisions is that antibiotics are prescribed for many diseases that may or may not require antibiotic treatment. Modeling antibiotic prescription for viral infections is contrary to traditional drug choice models, where the drugs are prescribed for assigned indications[5] To incorporate this feature, I let the physician's behavior differ by the type of disease based on whether an antibiotic prescription is justifiable. Regarding financial incentives, the bonus is paid at the end of the year and is a nonlinear function of performance. This might create dynamic incentives if physicians are forward-looking. The model provides two main insights. First, if physicians are forward-looking, their prescriptions should depend on the marginal cost of antibiotic prescription at the end of the year. As physicians form expectations over the end-of-year reward, the presence of uncertainty in future visits makes their response susceptible to the bonus design in a dynamic way. Second, given the consequences to the patients of the program, the response should differ by the necessity of antibiotic prescription. I test these predictions using changes in the bonus design in 2017 and 2018, which generates useful variation in terms of the marginal returns to prescriptions. I find that physicians' prescription rates decrease as the marginal cost of antibiotic prescriptions (that is, the foregone bonus) increases. They also respond to changes in incentives due to the nonlinear nature of the bonus function, in line with their expected performance. In the aggregate, the effect appears small in magnitude; however, this is partly a result of masked heterogeneity. To uncover the heterogeneity across diseases, I estimate a discrete choice model of antibiotic treatment for each disease using prescription-level data. The findings show that physicians do not change their behavior for diseases where an antibiotic treatment is almost *always* justifiable. This is an ensuring result, as the incentives otherwise would have been linked to unwanted consequences. They react moderately to diseases for which an antibiotic treatment is *sometimes* justifiable. These diseases could be of viral or bacterial origin, such as otitis and sinusitis. The

[4]Patient utility could reflect the real health benefit or the happiness of the patient, as patient pressure is shown to be a driving force of unnecessary antibiotic prescriptions (Kohut et al. (2020); Richards and Linder (2021)

[5]Antibiotic treatments are also different than the off-label use of medication, as in Dubois and Tunçel (2021), as there is no benefit of antibiotic prescription in the treatment of a viral infection.

largest impact is for diseases that are of viral causes. In 2019, the financial incentives decreased prescriptions for diseases that do not require antibiotics by 7% on average, yet there is significant heterogeneity across physicians. This number is 0.3% for diseases that almost *always* require antibiotics. While the effect of the program appears to be limited, it is also the case for the costs in the aggregate, but not necessarily at the prescription level. On average, physicians earned approximately €156 as a result of the program. However, the bonus is not contingent on showing improvement over years but is based on realized prescription rates [6]. Consequently, the cost of the program per saved antibiotic prescription cannot be dismissed as negligible. Each saved antibiotic prescription is rewarded with approximately €14, which is around 56% percent of the €25 fixed visit fees GP's earn.

Overall, the findings show that the financial incentives indeed impact physician behavior. However, the findings also emphasize the importance of thoughtful incentive design in terms of (i) the shape of the bonus function and (ii) the targeted pool of diseases. The P4P program represent a dynamic approach to physician decision-making. Physicians respond to discrete changes in the incentives created through the piece-wise linear feature of the bonus function. Therefore, conditional on a public budget, careful consideration of the bonus function could lead to better results. Furthermore, as the impact of financial incentives may vary depending on the necessity of antibiotic treatment for specific diseases, tailoring incentive structures to account for these variations may be required.

Literature This paper contributes to the literature on the effects of pay-for-performance in primary care. Pay-for-performance programs have been implemented in various regions around the globe as a method of incentivizing physicians, hospitals, or medical groups. This practice involves establishing specific criteria related to improving healthcare dimensions such as screening, prevention, and efficiency. Monetary rewards are attached to these criteria, which may differ in value depending on the specific design of the program. The literature thus far has used randomized controlled trials, quasi-experiments, and before-after studies with heterogeneous levels of validity concerns to evaluate different P4P programs. The evidence on P4P programs' success remains insufficient (Eijkenaar et al. (2013); Van Herck et al. (2010))[7]. This also applies to

[6]The program discourages performing worse than a reference year rate for physicians that are above a certain prescription rate (intermediary objective). For others, crossing this threshold (intermediary objective) would lead to a zero bonus. However, within the corresponding range, the bonus is independent of improvement realized and just a function of the antibiotic prescription rate, as explained in detail in 2

[7]Several works look at the effect of the P4P program in France. Sicsic and Franc (2017) Constantinou et al. (2017) uses data from CAPI (Contrat d'Amèlioration des Pratiques Individuelles) to study the impact on breast cancer screening in France. The first initiative for P4P in France is called Contracts for Improved Individual Practice (Contrat d'Amèlioration des Pratiques Individuelles - CAPI). CAPI worked as a trial period for a country-wide adoption of a P4P program. It took place between 2009 and 2011 and allowed individual contracts voluntarily for the incentive scheme based on some performance measures. In 2012, the program was restructured under the name Remuneration for Public Health Objectives(Rémunération sur Objectifs de Santé Publique - ROSP) to improve the quality of care, increase standardization across physicians, and reduce costs (Chevreul et al. (2015)). The authors observe which physicians signed the CAPI contract, and the time frame enables them to perform a difference-in-differences exercise to estimate the impact of the program on breast cancer screening. However, they do not provide any tests of parallel trends or identification of selection bias due to data limitations. They found a limited impact of the program. A before-after study of cervical cancer screening participation as a response to incentives introduced by ROSP in

antibiotic-related interventions (OECD (2023); Yoshikawa et al. (2021)). Among a few studies that consider the effect of P4P programs on antibiotic prescriptions, two stand out for their well-defined control groups. Mullen et al. (2010) studies the impact of P4P programs introduced in California in 2002 for a subset of medical groups, with an extension to cover more medical groups in 2003 (covering 60% of the revenue). They use difference-in-differences analysis to study the effectiveness of the program but also potential spillovers to unrewarded measures. Using the rate of preferred antibiotics in cases of bronchitis and pharyngitis as an outcome, they find a negative impact of this measure. The authors argue that the relative rewards in the design of the program put little emphasis on this criterion, which caused physicians to overlook and focus on more lucrative criteria. In Ellegård et al. (2018), the authors analyze P4P introduced in Sweden to increase the use of narrow-spectrum antibiotics instead of broad-spectrum ones in the treatment of children with respiratory tract infections. In their empirical design, they take advantage of decentralized healthcare systems at the county level. A total of 8 of 21 counties adopted P4P, which raises the questions of selection bias in adopting the P4P and also possible spillovers from treated to non-treated counties. They show that the incentives increased the share of narrow-spectrum antibiotic prescriptions in these areas. I add to the existing literature by testing the hypothesis that if the incentives are successful, they should affect physicians more prone to antibiotic prescriptions that are not *always* justifiable by using regional variation in viral diseases.

I also add to the literature on the effect of dynamic incentives created by the design specifics of the bonus. To the best of my knowledge, this paper is one of the few to study the forward-looking behavior of physicians in such a setting[8]. There are three possibilities if physicians react to financial incentives: (i) they learn from them and change their behavior consistently, which would preclude dynamic incentives, (ii) they purely react to the bonuses, so if the bonus function is poorly designed, one might miss on further improvement, and (iii) they learn but still react to incentives. If they are reacting to bonuses, how they react is important for the bonus design since myopic and forward-looking behavior would have different implications for physicians' responses. One example in the literature that signals forward-looking behavior is Dowd et al. (2013). In

2012 was also performed. They identify a positive effect on screening participation compared to the rate in 2011. In a recent study, Sanchez et al. (2023) look into temporal variation in the ROSP scores of physician in the Grand-Est region of France from 2017 to 2020. The authors point out to overall improvement in the scores and differences across the rural and urban parts of the region.

[8]The dynamic incentives are also studied in the context of pay-for-performance programs for sales force compensation (Chung et al. (2014); Misra and Nair (2011)). However, it is possible to consider this problem in reverse, i.e., the bonus not earned by each prescription of antibiotics is essentially the price that the physician pays for that prescription. Therefore, the problem at hand parallels dynamic incentives due to nonlinear pricing contracts that have been studied in IO literature in the context of health insurance (Aron-Dine et al. (2015); Brot-Goldberg et al. (2017); Dalton et al. (2020); Einav et al. (2015); Klein et al. (2022)) and cellular and data usage pricing (Grubb and Osborne (2015); Nevo et al. (2016)). An important distinction is that there is no real spot price since the bonus is paid at the beginning of the following year. This effectively forces physicians to consider expected end-of-year performance levels and returns to bonuses for this expectation. I provide evidence of this behavior using a change in the kink points of the bonus function in 2017. Since physicians can observe their (current) performance, the uncertainty is due to the incoming disease and patient profiles for the rest of the year at a given time. There are two sources of information; first, most of the software used to digitize prescriptions also keeps track of the P4P scores. Additionally, doctors receive official updates in every quarter through their Ameli accounts.

contrast to my setting, the incentives are provided at the group practice level and rewarded as all-or-nothing. The practices that are far below the target appear to prescribe as if the incentives do not exist. Rosenthal et al. (2005) studies a different setting in which the rewards are at the physician group level again but the bonus is not all-or-nothing but rather a payment per physician who reaches a target. The authors find that for the criterion for which performance improved, it was improved by the group whose baseline performance was the worst. Although these studies investigate physician-group-level incentives, they nonetheless signal how the design can matter - going from a binary bonus to a linear one based on the aggregate achievement rates.

The rest of this paper is organized as follows. Section 2 introduces the pay-for-performance program in the French healthcare context. It also presents data and provides descriptive statistics regarding physician prescription behavior. Section 3 explores how the incentives target the discrepancy in socially desired levels of antibiotic prescriptions and the observed levels and shows the effectiveness of incentives using heterogeneity in the disease profiles physicians treat. Section 4 introduces the theoretical model and details the predictions. Section 5 empirically tests these predictions and states the findings. Finally, Section 6 concludes.

2 Context and data

The healthcare system in France is built upon universal healthcare coverage. General practitioners (GP) serve as the gate-keepers in the system and act as the primary care points (Or et al. (2023)). Each patient is assigned to a general practitioner, and this physician acts as the first point of contact. GPs earn a fixed fee per visit (€25) as their main source of income. As I detail below, they also receive payments based on their practices with the introduction of the P4P system.

2.1 Pay-for-performance in French healthcare

In 2012, bonus payments to physicians were introduced at the national level through the Remuneration for Public Health Objectives (Rémunération sur Objectifs de Santé Publique - ROSP) program. This performance-based payment program aims to improve the quality of care, increased the standardization of practice across physicians, and reduce costs (Chevreul et al. (2015))[9]. The bonus scheme covers general practitioners, cardiologists, endocrinologists and general practitioners for pediatric patients working in private practice. All physicians are enrolled in the bonus scheme unless they choose to opt out. In 2018,

[9]The trial for bonus payments for physicians in France is called Contracts for Improved Individual Practice (Contrat d'Amèlioration des Pratiques Individuelles - CAPI). CAPI worked as a trial period for a country-wide adoption of a P4P program. It took place between 2009 and 2011 and allowed individual contracts voluntarily for the incentive scheme based on some performance measures.

approximately €250 million was paid as performance bonuses to 50785 general practitioners[10][11]. In this project, I focus on general practitioners who are responsible for most of the antibiotic prescriptions, with 70% of all antibiotics in the outpatient setting in 2015)[12].

The P4P program for GPs is based on 29 clinical practice indicators/criteria in 2019. The set of indicators and the bonus design were updated in 2017 and 2018[13]. The health authority defines the reward function for the indicators. The reward function is a function of performance, which is calculated as a percentage depending on the objective. Examples include the percentage of patients who are vaccinated for flu and the percentage of patients who are prescribed antibiotics. Through this reward function, physicians collect points for each indicator based on their performance. For instance, the fewer antibiotics they prescribe, the higher the bonus they earn for that criterion. At the end of the year, the physicians are paid based on the total number of points they collected and the number of patients who registered them as their primary GP. The indicators belong to three main groups: monitoring of chronic pathologies (8 indicators, in total 220 points), prevention (12 indicators, in total 390 points) and efficiency (9 indicators, in total 330 points)[14].

Performance definitions related to antibiotic prescriptions This paper focuses on two criteria that target better use of antibiotics. The first criterion is "Overall Antibiotics Criterion" (OAC). OAC has been a part of the program since 2012 and concerns the (unconditional) rate at which physicians prescribe antibiotics. For OAC, performance is defined as the number of antibiotic treatments per 100 (low-risk) patients. The program excludes patients who are older than 65 or suffer from long-term illnesses (ALD) as a risk adjustment. "At-risk Antibiotics Criterion" (AAC) was introduced in 2017. For AAC, performance is defined as the share of patients who received at-risk antibiotics among patients treated with antibiotics. The at-risk antibiotics include amoxicillin and clavulanic acid, 3rd and 4th Gen. cephalosporins and fluoroquinolones. The selection is justified by these antibiotics being broad-spectrum antibiotics that are still effective against multi-resistant bacteria, and the resistance against them has shown a concerning increase. Note that for both criteria, the reward is a decreasing function of the rate as the program encourages fewer prescriptions, especially of "at-risk" antibiotics. Therefore, while the OAC is designed to reduce the rate of antibiotic prescriptions, primarily among low-risk patients, without distinguishing between different

[10]In 2018, the number of generalists was 102000, with 57% of them working in private practice, meaning that the share of physicians who actively opt-out is quite small.

[11]Payment in 2018 retrieved from `https://www.hcsp.fr/Explore.cgi/Telecharger?NomFichier=ad1080407.pdf` on 25/10/2023. Number of general practitioners in France Retrieved from `https://drees.solidarites-sante.gouv.fr/sites/default/files/2020-08/dossier_presse_demographie.rdf` on 25/10/2023

[12]`https://www.omedit-grand-est.ars.sante.fr/system/files/2017-08/ANSM-rapport-antibio_2016_bd2.pdf`

[13]The set of indicators and the rules were updated during the sample period according to Article 27 and Annex 15 of the medical convention of 25 August 2016 and amendment 6 of the medical convention published in Official Journal on 10 August 2018.

[14]Monitoring of chronic pathologies: diabetes, arterial hypertension, cardiovascular risk, prevention: flu vaccination, cancer screening, iatrogenicity, antibiotic therapy, addictive behaviors, efficiency: prescription in the generic directory, biosimilars, efficiency factors.

types of antibiotics, the AAC focuses on decreasing the proportion of a specific category of antibiotics, which includes certain broad-spectrum antibiotics, among all antibiotic prescriptions.

2.1.1 Bonus function

Let R denote the reward per criterion, b^c the bonus coefficient (which is a function of performance), p^c the points collected per criterion, $\bar{p}$ the maximum points per criterion and N the number of patients who declare the physician as their attending physician. The points are calculated at the criterion level through the following formula:

$$p^c = b^c \times \bar{p}$$

, where the bonus coefficient is set by the program as a piece-wise linear function of performance, as I detail in the next section. The maximum points possible, $\bar{p}$, is set as 35 for both criteria. Then, for each point, the physician receives a bonus with a physician-specific conversion rate:

$$R = p^c \times \frac{N}{800} \times 7 \text{ (€)}.$$

Bonus coefficient The complex object in the calculation is the bonus coefficient. The bonus coefficient is a piece-wise linear function of physician performance at the end of the year, Perf_t, and the function itself depends on three components: (i) physician's performance at the end of the reference year, which defines the point where physicians start to obtain a nonzero bonus, (ii) target rate (T), which is set by the program and grants physicians a full bonus upon reaching it, and (iii) intermediary objective (IO), which is set by the program and grants physicians 30% of the bonus upon reaching it [15][16]. Let P_{it} denote the performance of physician i in year t and r_i the performance in the reference year. The bonus coefficient function is written as follows: [17]:

$$b^c(\text{P}_{it}) = \begin{cases} 0 & \text{if } \text{r}_i < \text{P}_{it} \\ 0.3 \cdot \frac{\text{r}_i - \text{P}_{it}}{\text{r}_i - \text{IO}} & \text{if IO} < \text{P}_{it} < \text{r}_i \\ 0.3 + 0.7 \cdot \frac{\text{IO} - \text{P}_{it}}{\text{IO} - \text{T}} & \text{if T} < \text{P}_{it} < \text{IO} \\ 1 & \text{if } \text{P}_{it} < \text{T} \end{cases} \tag{1}$$

[15] The achievement rate for meeting the intermediary objective was initially 50% and was updated in 2018.

[16] According to the guidelines provided by HAS, the intermediate objective and target are defined based on (i) public health objectives, when these are specified in the literature (prevention section in particular) and (ii) the distribution by deciles of physicians observed for each indicator. In the second case, the intermediate objective is set at the 3rd decile (7) (for an increasing indicator) or the 7th decile (3) (for a decreasing indicator) and the target objective is set at the 8th decile (9) (for an increasing indicator) or the 2nd decile (1) (for a decreasing indicator). The changes in the design implemented in 2017 and 2018 include the deciles used. There is no further transparency on how they are set.

[17] Note that the formula applies when the decrease in the specified rate (i.e., performance) is associated with higher rewards. Antibiotic-related criteria are examples of this case, as lower rates are better.

Table 1 provides a summary of the bonus design parameters and the changes across years.

Table 1: Bonus coefficient function design parameters across years

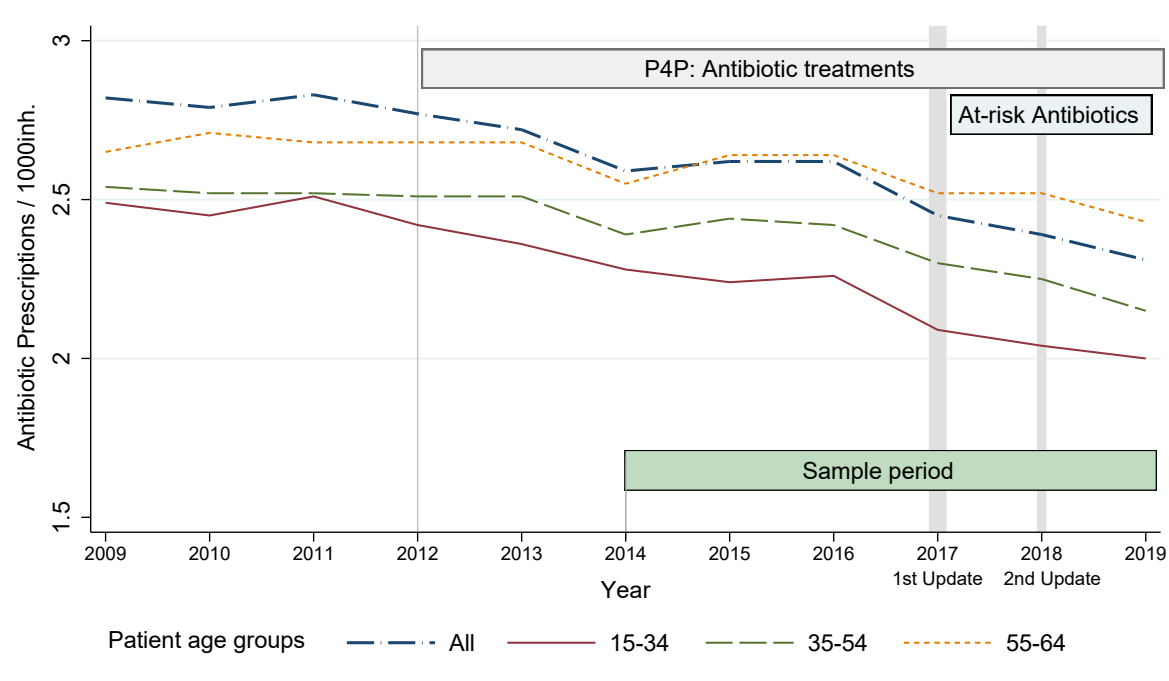
Period	Criteria	Kink Points	
2012 - 2016:	OAC: # antibiotic treatments per 100 low-risk patients	T	**37%**
		IO	40%
2017		T	**14%**
		IO	25%
2018 - 2019		T	**20%**
		IO	45%
2017	AAC: Share of patients received at-risk antibiotics	T	**27%**
		IO	36%
2018 - 2019		T	**32%**
		IO	52%

Notes: The bonus is a piece-wise linear function of performance, as presented in 1. This table presents the kink points of the function, T: Target rate, IO: Intermediary objective.

The alterations presented in Table 1 introduce valuable variations in the performance standards required to earn higher rewards. Before 2017, the OAC had two thresholds: the target rate (T) and the intermediary objective (IO), which were relatively close. In 2017, both thresholds were adjusted to become more stringent. Starting in 2018, the thresholds were again updated, creating a 25-percentage point performance difference gap, where the bonus function followed a linear pattern with a common slope for all physicians. Another significant change in 2017 was the introduction of a new criterion, the AAC, into the program. The AAC also underwent an update in 2018.

Figure 1 illustrates the aggregate use of antibiotics per 10,000 inhabitants across various age groups. While this measure does not directly map to the performance criteria defined above, it nonetheless serves as a proxy for antibiotic usage. We can observe a decrease immediately after 2012, which gradually levels out until 2017, when stricter requirements were needed to attain the same level of bonus. This period also marked the inclusion of the AAC in the program.

Figure 1: Trend in antibiotics prescriptions per 1000 inhabitant by age groups



Notes: The plot uses aggregate data from Sante publique France - SNDS - INSEE. The data provide antibiotic prescriptions per 1000 inhabitants for different age groups and all age groups across France.

In the following section, I explain the data, covering the years from 2014 to 2019, encompassing the two bonus design changes, as depicted in Figure 1. In Sections 3 and 5, I provide a detailed description of how I utilize these changes to assess the effectiveness of the incentives.

2.2 Data

2.2.1 Prescriptions data

I use prescription-level data from a representative sample of physicians from 2014 and 2019 (CEGEDIM). For more than 2000 GPs, data record all prescriptions in detail as well as information about the visit and patient and physician demographics. To have a balanced sample, I keep the physicians present in all years [18] Thus, the analysis includes 1116 physicians.

On the physician side, I observe the age, gender and region of operation in France. On the patient side, the data include age, gender, and chronic diseases. For each appointment leading to a prescription, I observe the diagnosis code (with International Classification of Diseases (ICD) codes), molecule (Anatomical Therapeutic Chemical Classification), product code (Code Identifiant de la Présentation - CIP 13), manufacturer, frequency/dosage of use, and co-prescriptions. Summary statistics are presented in Table 2.

[18] I also exclude the first and last percentile of physicians in terms of the number of patient visits, the physicians who operate less than 10 months in a given year, and those who prescribed fewer than 50 antibiotics.

Table 2: Summary Statistics 2014-2019

Physicians	Mean	SD	Min.	Max
Age	55.3	8.4	28	80
Female	0.3	0.5	0	1
Visits per year	3343.6	1401.4	297	12230
% of visits - *Always* bacterial diseases	2.2	1.0	0.0	7.6
% of visits - *Sometimes* bacterial diseases	14.6	4.7	0.7	33.9
Patients per year	1139.6	439.2	134	2938
Avg. (base) patient chronic cond.	0.3	0.1	0.0	1.2
Avg. patient age	53.7	5.3	36.6	71.9
% of patients of 16-65 years old	69.3	11.5	27.2	97.5
Patients (with at least one antibiotic prescription)	Mean	SD	Min.	Max
Age	49.0	19.0	16	120
Female	0.6	0.5	0	1
Nb. excluded chronic diseases	0.6	0.9	0	7
Nb. base chronic diseases	0.2	0.6	0	7
Visits per year	4.8	4.0	1	139
Antibiotic treatments per year	2.0	1.6	1	32

Notes: Number of (distinct) physicians: 1116, number of (distinct) patients: 1394641

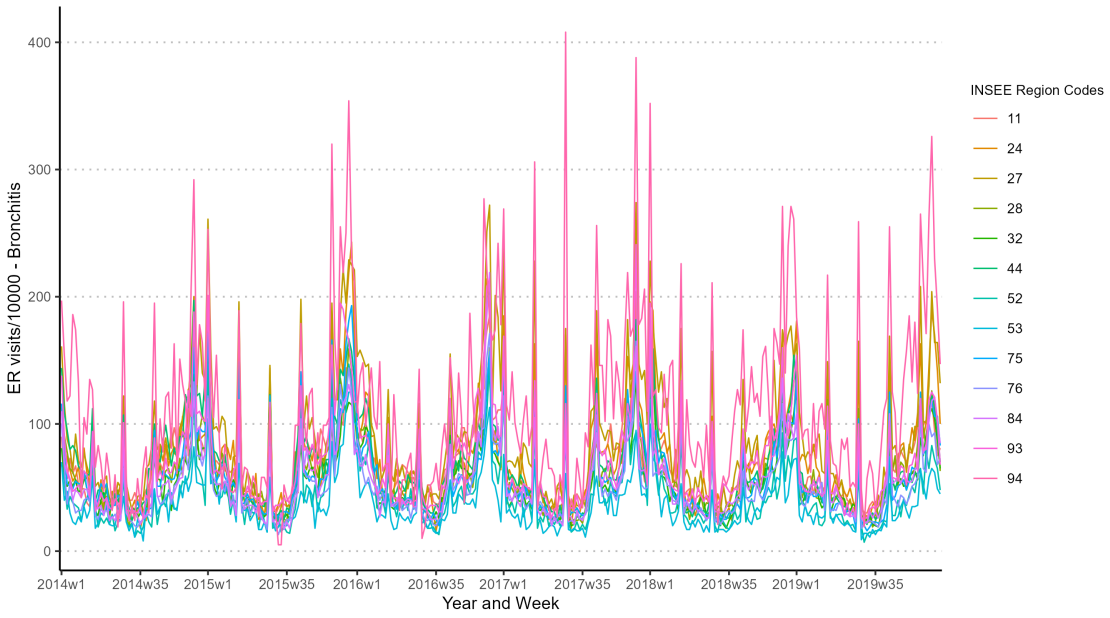
Physicians exhibit substantial differences in the characteristics of their patient populations, both in terms of patient attributes and the size of their practices. Furthermore, they vary in the types of medical conditions they address among their patients. Notably, the diversity in the prevalence of diseases based on their bacterial or viral nature presents a substantial variation. Physicians who primarily treat conditions that necessitate antibiotic usage face challenges in reducing antibiotic prescriptions and complying with the OAC. However, they have the opportunity to enhance their performance concerning the AAC by decreasing the overuse of specific high-risk antibiotics. Conversely, physicians handling a higher volume of patients with viral illnesses and prescribing antibiotics have a greater potential to earn bonuses through the OAC. The prescription of antibiotics for viral conditions also introduces temporal patterns, creating seasonality in their practices. This variation will guide the reduced form analysis in Section 3, which speaks to the effectiveness of the incentives.

2.2.2 Regional epidemiological indicators - Géodes

I use publicly available data on public health indicators constructed by the public health authority in France. The data are constructed using various sources, including surveillance systems, epidemiological surveys and

the National Health Data System. The variables vary in the granularity of the geographical region and also in terms of the time frame. In Section 3, I rely on two indicators: the rate of visits to emergency rooms for acute bronchitis (per 10,000 visits) and the rate of emergency visits for flu (per 10,000 visits), which that are provided weekly at the regional level[19]. Figure 2 presents the data for acute bronchitis.

Figure 2: Rate of visits to emergency rooms for acute bronchitis per 10,000 visits by regions of France



2.2.3 What are antibiotics prescribed for?

Estimates for unnecessary antibiotic prescriptions range from 30% in the US (Fleming-Dutra et al. (2016)) to 50% in France (Carlet and Le Coz (2015)). Antibiotic prescription stands as the "easier" option for several reasons, including patient satisfaction and pressure, time constraints (limited time to see patients, diagnose their illnesses, formulate a treatment plan, and avoid lengthy explanations of why the drugs are not needed), decision fatigue, and uncertain diagnoses [20].

A physician prescribing antibiotics to a patient who does not need them leads to several undesirable outcomes from a global health perspective, which can be assimilated to a type 1 error in the traditional testing setup[21]. However, the observed prescriptions could also be a result of mitigating the type 2 error, hence failing to prescribe antibiotics to a patient who genuinely needs them. Assuming the diagnosis can be observed with certainty by the physician, the objective of the policy is to minimize the occurrence of

[19]The data are from an emergency surveillance organization OSCOUR - Organisation de la surveillance coordonnée des urgences.

[20]A non-exhaustive list of evidence includes Cole (2014); Guillemot et al. (1998); Kohut et al. (2020); Lévin et al. (2019); Richards and Linder (2021)

[21]These include overuse of antibiotics, contributing to antibiotic resistance, potential side effects without any therapeutic benefit, and increased healthcare costs.

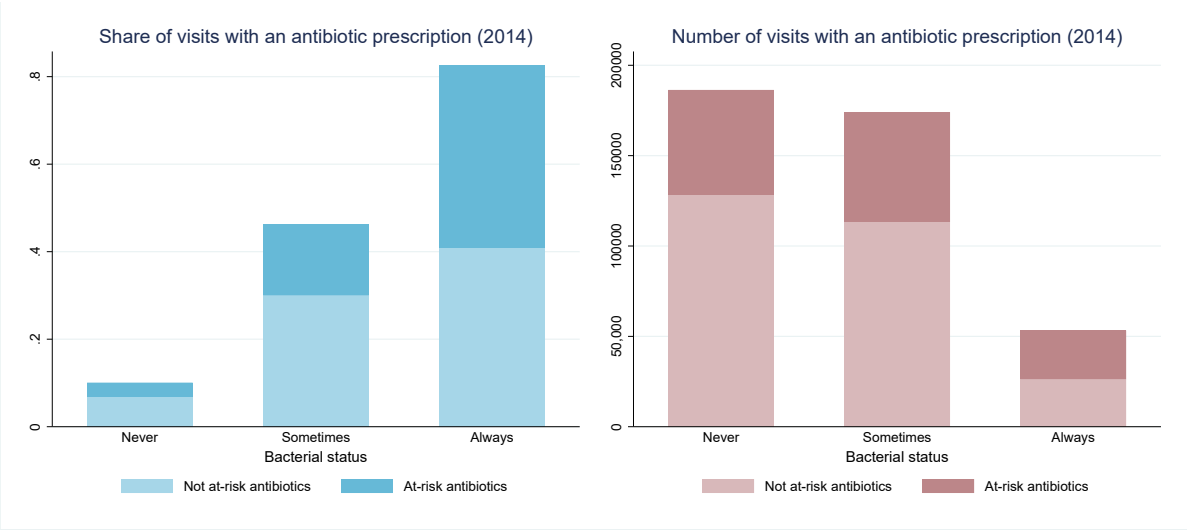
the two types of error, accounting for the private incentives of the physician. Under diagnostic uncertainty, the probability associated with each type of error would remain positive. Because of how the incentives are structured, there is no differentiation between the different types of errors. The observed action by the policy maker that can be contracted upon is the antibiotic prescription rate. While this is expected to decrease type 1 errors, it could also increase type 2 errors. In the following analysis, I assume that the reported diagnosis codes match the final decision of the physician. Therefore, based on the diagnosis at hand, I classify antibiotic prescriptions as type 1 errors if the cause is a viral infection. In Section 4, I provide a framework to understand the effect of policy on different types of errors and show that the impact of the policy is higher for diseases for which we observe type 1 errors. Moreover, I find no impact of the policy increasing type 2 errors for diseases that are of bacterial origin.

In this paper, I use the classification of diseases suggested by Chua et al. (2019) based on whether antibiotic treatment is justifiable. The authors provide an extensive list of diseases identified at the ICD-10 level, each attached to a status $\{$*never* (justifiable), *sometimes* (justifiable), *always* (justifiable)$\}$[22]. The differentiation is based on whether a disease is associated with a bacterial infection. Examples include pneumonia or urinary tract infection for *always*, acute sinusitis or acute otitis media for *sometimes*, and acute upper respiratory tract infection, acute bronchitis, which are almost surely viral infections or noninfectious conditions, for *never*. The data include the diagnosis codes for the treatment provided by the physician at the time of prescription. I use *always* (bacterial), *sometimes* (bacterial) disease and *never* (bacterial) disease to indicate the group to which a disease belongs. Figure 3 shows the shares and numbers of antibiotic treatments (I divide antibiotics into two groups, as defined by the bonus program: at-risk antibiotics and not at-risk antibiotics) by each group in 2014 for patients covered by OAC[23]. The physicians in the sample prescribed antibiotic treatments 80% of the time to diseases they concluded to be in the *always* group. This rate is around 40% for the *sometimes* group and 10% for the *never* group. Although low in the share of antibiotic treatments, the prescriptions for the *never* group sum to around 47% of the total antibiotics prescribed.

[22] I group diseases at a higher level and assign the status based on the majority of the cases seen in the data.

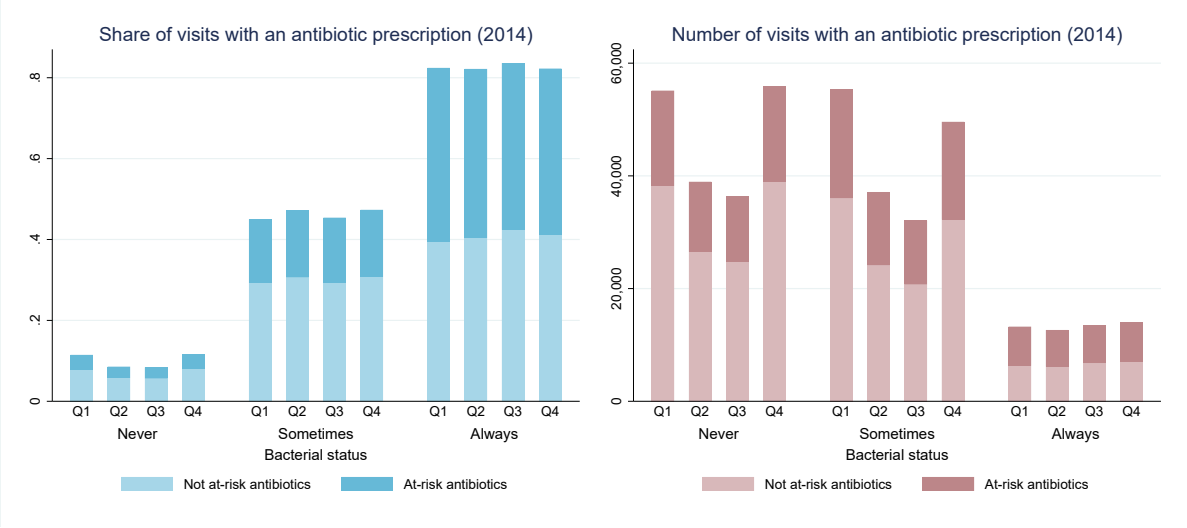
[23] I choose to report the values for 2014 since it is the year that is likely to be least affected by the bonus scheme as it is the earliest year in my sample.

Figure 3: Antibiotic treatment by group of diseases



The observed pattern of antibiotic prescription by disease group also presents seasonality. Figure 4 shows the prescriptions for each group across quarters. The seasonality could be a result of type 1 errors and shown to be causing fluctuations in resistance rates Sun et al. (2012).

Figure 4: Antibiotic treatment by group of diseases and seasonality



The diseases within each group have varying rates of antibiotic treatment, which is most pronounced for the *never* diseases. In Table 3, I report the number and share of antibiotic treatments for the top 5 diseases in terms of antibiotic prescription rate within each group. The most striking is the treatment for bronchitis, which receives an antibiotic prescription 70% of the time.

Table 3: Top 5 diagnoses (by antibiotic treatment rate) by group (2014)

Disease Group	ICD Name	Antibiotic treatment	
		Count	Share
Never			
	Acute bronchitis	20950	72.61 %
	Bronchitis, not specified as acute or chronic	22077	71.86 %
	Pain associated with micturition	2544	61.63 %
	Other symptoms and signs involving the circulatory and respiratory systems	670	52.39 %
	Other disorders of teeth and supporting structures	1627	48.58 %
	Others	138430	7.77 %
Sometimes			
	Acute sinusitis	31305	79.13 %
	Bacterial infection of unspecified site	710	77.85 %
	Inflammatory diseases of prostate	1448	76.13 %
	Suppurative and unspecified otitis media	23403	75.52 %
	Other bacterial agents as the cause of diseases classified elsewhere	1330	62.97 %
	Others	115822	38.40 %
Always			
	Other disorders of the urinary system	10838	90.5 %
	Cystitis	27681	90.25 %
	Acute pyelonephritis	1118	85.21 %
	Urethritis and urethral syndrome	968	83.95 %
	Diseases of pulp and periapical tissues	4919	80.91 %
	Others	7684	58.04 %

Notes: Number of antibiotic prescriptions is the number of visits where at least one antibiotic was prescribed in 2014. The share is the number of antibiotic prescriptions divided by the total number of visits.

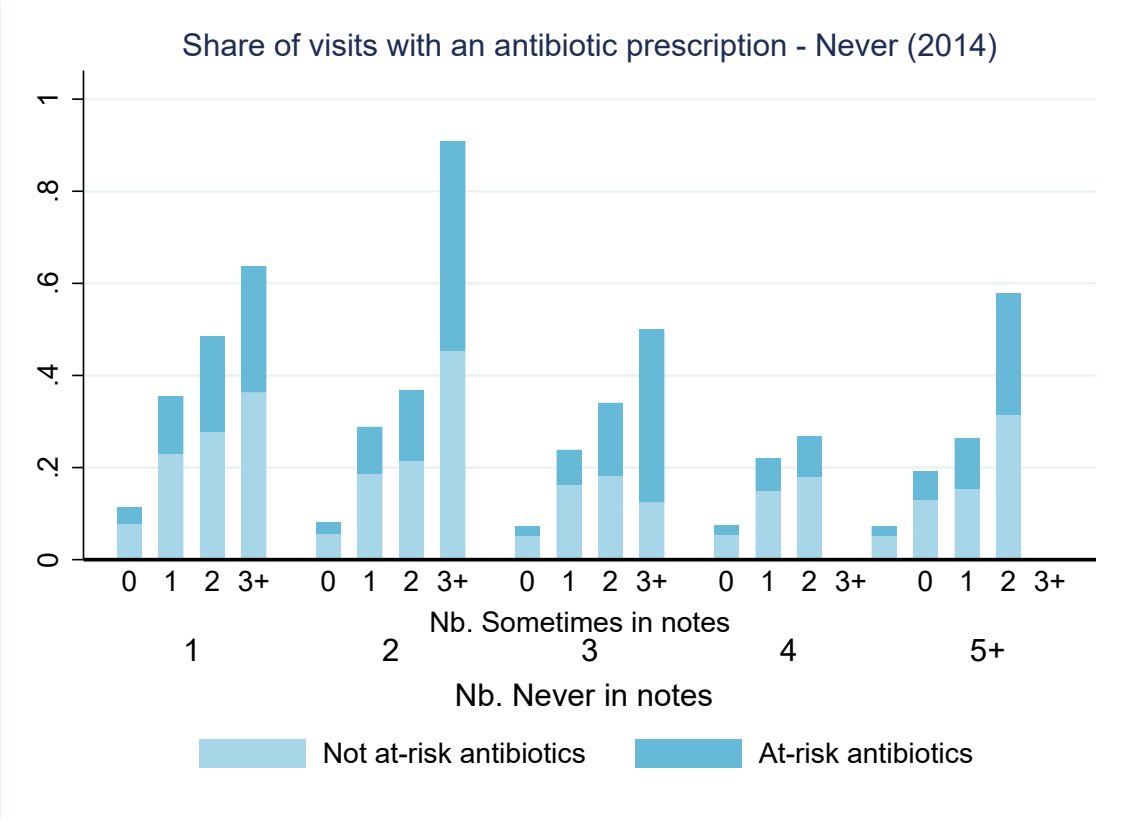
Assuming there is no diagnostic uncertainty, the high share of antibiotic treatments for *never* diseases raises a concern in the fight against AMR and motivates the actions taken to reduce externalities. However, the data show evidence of diagnostic uncertainty, which I explore next.

Diagnostic uncertainty When there is diagnostic uncertainty, there is not necessarily a one-to-one mapping between the diagnostic codes in the data and the justifiability of antibiotics if physicians are risk-averse against type 2 error. To explore this situation, I use information on the stated reasons for the appointment. These are the notes of physicians in the form of diagnostic codes and provide valuable information on diagnostic uncertainty. However, there are differences in prescription behavior across the stated reasons for prescriptions.

To show this, I count the number of diseases in the initial contact record that belong to each group. Then, for each of the reported diagnosis groups for the prescription, I plot the share of antibiotic treatments for different combinations of disease groups in the initial contact notes. Figure 5 provides the share of

antibiotic treatments for which the disease stated for the prescription is a *never* justifiable disease. The shares are provided for different categories based on what is in the contact diagnosis information. The graph shows that if there is a mention of a *sometimes* disease in the contact diagnosis, the chances of antibiotic prescription increase. As the number of *never* diseases increases, the uncertainty resolves in the direction of non-antibiotic treatment, whereas as the number of *sometimes* diseases increases, the uncertainty resolves in the direction of antibiotic treatment being a justifiable option. The prescription rates therefore are low when there is no mention of *sometimes* (or always) diseases and increase with *sometimes* diseases. This increase is less when the number of *never* diseases is higher.

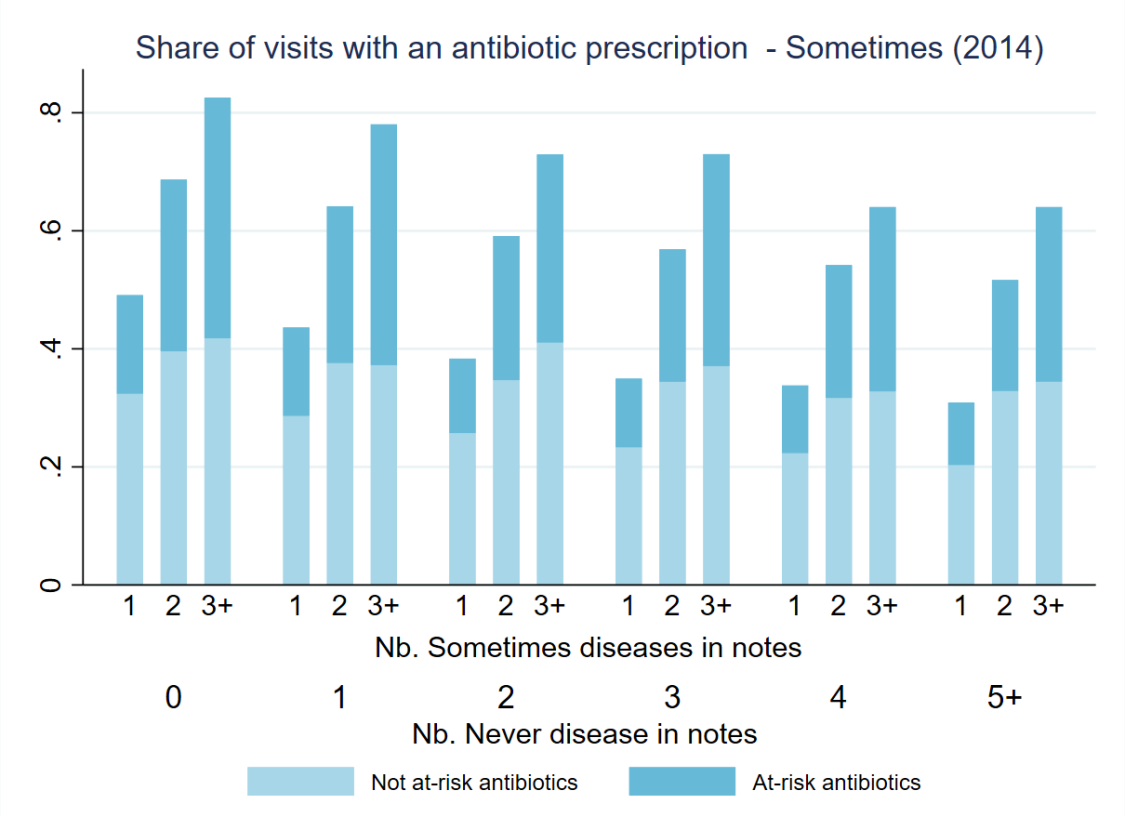
Figure 5: Antibiotic treatment rate by the number of different disease groups in notes when the prescription diagnosis belongs to the *never* group



Notes: The few observations where there is an *always* group disease in the notes are excluded.

Similar patterns are also observed for diseases that belong to the *sometimes* group, Figure 6.

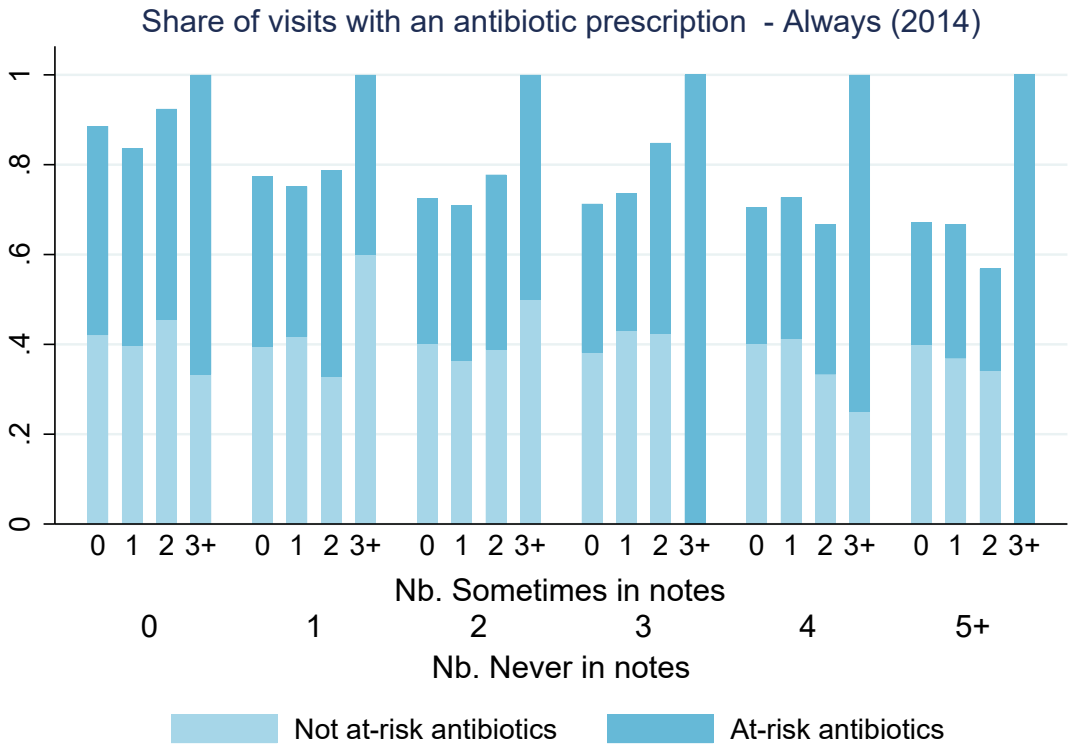
Figure 6: Antibiotic treatment rate by the number of disease groups in notes when the prescription diagnosis belongs to the *sometimes* group



Notes: The observations where there is an *always* group disease in the notes are excluded.

For the *always* group, on the other hand, the number of *sometimes* diseases does not carry as much information as the number of *never* diseases. This could be due to the fact that the *always* group is a stronger signal of antibiotic prescription requirement; hence, *sometimes* group diseases do not add informational value.

Figure 7: Antibiotic treatment rate by the number of disease groups in notes when the prescription diagnosis belongs to the *always* group

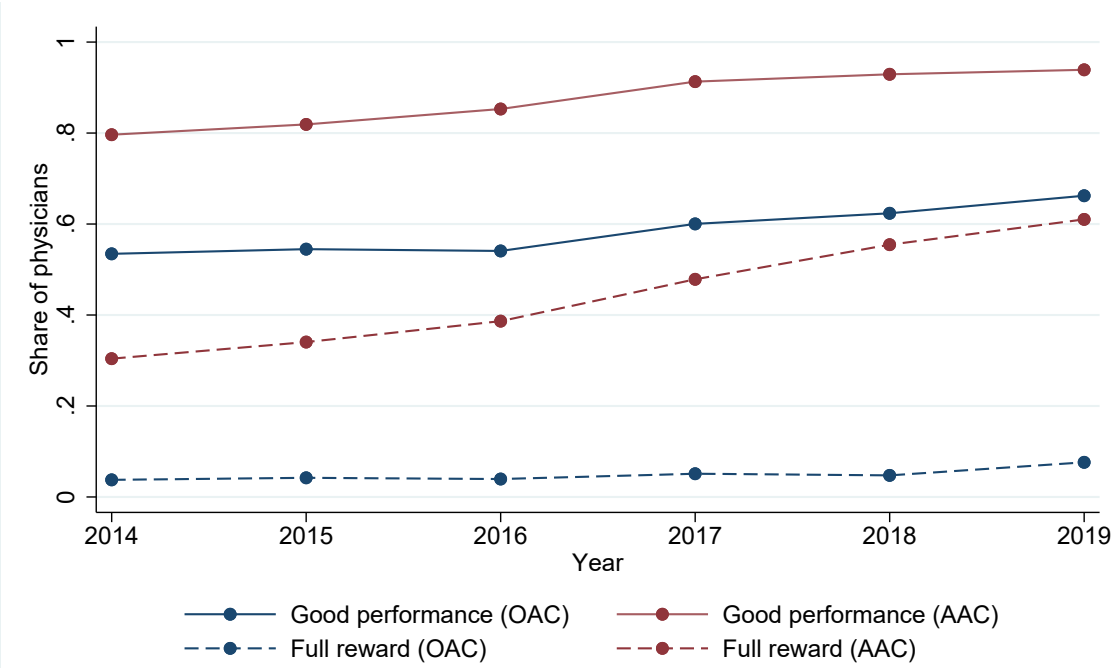


In the analysis at the disease level in Section 5, I control for the uncertainty using the content of the initial contact notes. However, the stated reason for antibiotic prescription is in line with the prescription behavior of the physician, providing support for the assumption that the diagnosis on the prescription aligns with what the physician thinks in terms of the cause of the disease. For any combination of the number of *sometimes* and *never* diseases, the ordering of the antibiotic prescription rate follows *always* (Figure 7), *sometimes* (Figure 6) and *never* (Figure 5).

2.2.4 Trends in performance

Improvements are present already in the aggregate measure. Figure 8 presents the share of physicians who perform better than the intermediary objective (good performance) and better than the target rate (full bonus) based on 2019 values. For the OAC, a shift across performance groups is observed upon the introduction of more stringent requirements from 2017 onward. For the AAC, physicians in the sample perform quite well already, and slight improvement is already present before the addition of the criterion to the program in 2017.

Figure 8: Evolution of performance groups across years



Notes: Good performance is defined as rates that are below the intermediary objective in 2019 (45% for overall antibiotics criterion and 52% for at-risk antibiotics criterion), and full bonus is when the physicians are below the target rate (20% for overall antibiotics criterion and 32% for at-risk antibiotics criterion).

3 Impact of stricter requirements on antibiotic prescriptions

In this section, I provide evidence to support the effectiveness of incentives. I do this by leveraging changes in the bonus function's design in 2017 and considering the heterogeneity in the diseases that physicians treat. Notably, in 2017, the existing pay-for-performance program underwent restructuring, affecting the calculation of bonuses and the criteria by which physicians are assessed, as detailed in 1. The thresholds, target and intermediary objective, of the bonus function, changed for all criteria in 2017 and also in 2018[24]. Second, a new criterion was added to the program, i.e., AAC. These changes offer a unique opportunity to illustrate how incentives can effectively enhance performance.

The variation exploited in this section is motivated by the use of antibiotics presented in Figure 3. Two observations stand out. First, a large proportion of antibiotics is prescribed to diseases that belong to the *never* category, followed by the *sometimes* category. Therefore, to the extent that these prescriptions are the results of type 1 errors, the margin of response to the incentives is whether to prescribe or not. Furthermore, the highest rate of prescription of at-risk antibiotics is for the group of diseases where an

[24]The change in 2018 was announced by amendment 6 of the medical convention published in the Official Journal on 10 August 2018. As the new rules were applied to physicians for the payment of 2018 as well, there was also some compensation for the possible loss. Therefore, the results of 2018 should be treated with caution.

antibiotic prescription is *always* bacterial. Consequently, for these diseases, a physician's choice primarily revolves around the selection of the appropriate antibiotic, i.e., whether an "at-risk" antibiotic is necessary, rather than the decision of necessity of antibiotic therapy. To show that physicians react to the new criterion that concerns at-risk antibiotics, I exploit the variation in incoming cases of *always* class diseases. In the following, I explain the empirical strategy and present the results for the two criteria.

3.1 Overall antibiotic prescription criterion

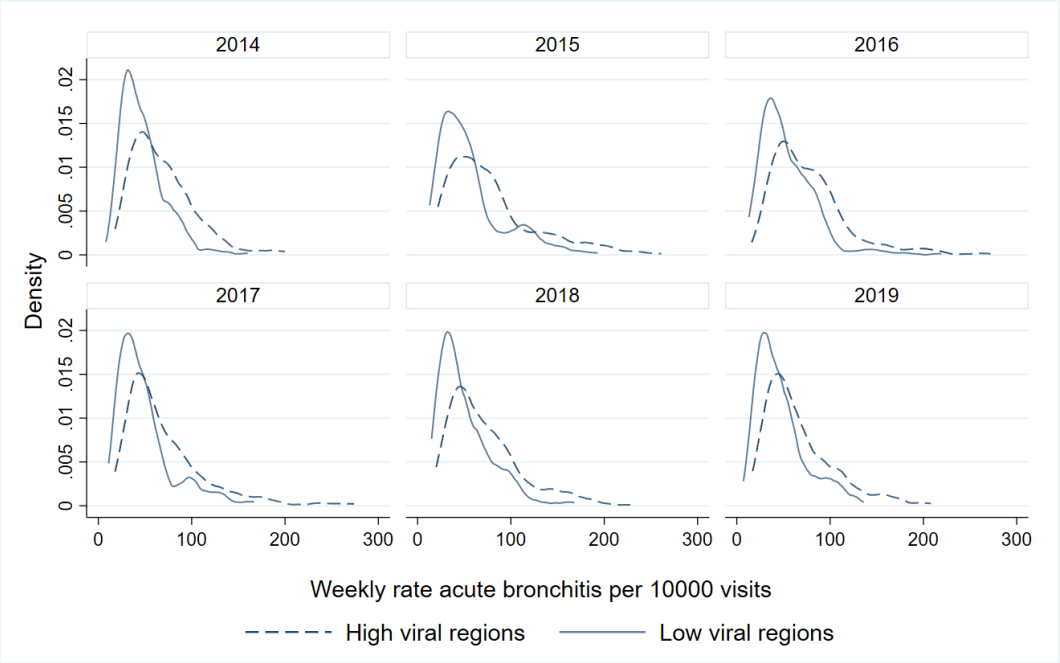
Physicians practice in diverse geographical areas, marked by unique characteristics, including variations in viral infection rates. As previously highlighted, viral infections, including common conditions such as bronchitis (as outlined in Table 3), contribute substantially to the overall volume of antibiotic prescriptions. I exploit this geographical variation and viral infection rates to evaluate the effectiveness of the incentives.

To do so, I group physicians into "pseudo" control and treatment categories based on the prevalence of viral infections in their respective regions of practice. Assuming altruistic physicians and given that antibiotics are valuable against bacterial infections only if incentives are working, the effect should be stronger for physicians who face more viral cases that give rise to type 1 errors in prescriptions. The rationale is that reducing antibiotic prescriptions for diseases of viral origin poses a lesser risk to patient health compared to cases involving bacterial infections; hence, they are less costly. The results show that physicians operating in regions with a higher prevalence of viral infections experienced a more substantial reduction in antibiotic prescription rates in response to the stricter requirements implemented from 2017 onward.

High and low viral infection prone regions France is divided into twelve metropolitan administrative regions (excluding Corsica). Each physician's region of operation is observed in the data. I complement the prescriptions data with the rate of visits to emergency rooms for acute bronchitis per 10,000 visits from OSCOUR network, as presented in Section 2.2.2. To assign physicians to high and low viral infection-prone regions, I use the following strategy. Using the rate of visits to emergency rooms each week for acute bronchitis per 10,000 visits, I rank regions from 1 to 12 in decreasing order. Using ranks instead of levels alleviates the problem of sensitivity to extreme values and provides a normalization that enables the analysis to focus on relative differences rather than absolute ones. Each week, the higher the rank, the higher the ER visit rates. Then, to construct a measure to allow comparison across years, I take the median rank of each region within a year and take the mean over these median ranks across years for each region. This allows for comparisons using an aggregated measure that provides a summary of each region's performance. These averages across regions are presented in Figure 23. From here on, I define a region as a high viral infection region if the average median rank of the region is higher than 6. The distribution of weekly ER rates for

high and low viral infection regions across years is presented in Figure 9.

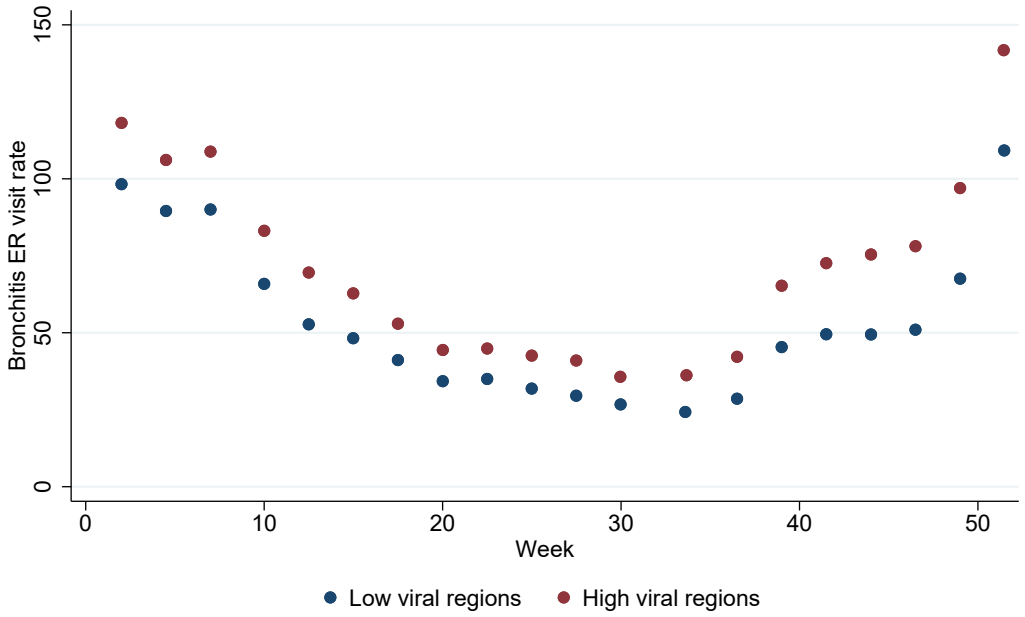
Figure 9: Distribution of weekly ER rates for the high and low viral infection prone regions



Notes: The distribution is across weeks and regions for each year.

The comparison is based on the aggregate ranks derived from the weekly rates. To see that the comparison carries to the weekly level, Figure 10 presents the unconditional means of weekly rates across weeks for the high and low viral infection regions.

Figure 10: Average weekly rates for high and low viral infection prone regions



Notes: Each dot is an (unconditional) average rate across years and regions in the given week.

Empirical analysis and results As a physician operates in a single region, the assignment of regions based on how viral infection-prone they are is equivalent to assigning physicians to the control and treatment groups. For physician i, I define the outcome variable as the monthly prescription rate. This allows me to have granularity in the observations and control for factors that influence the prescription rate of the physicians in a more flexible way[25].

I estimate the following linear probability model, where Y_{it} is the prescription rate of physician i at month t[26].

$$Y_{it} = \alpha H_i + \sum_{y' \in \{2015,\ldots,2019\}} \beta^{y'} \cdot H_i \cdot \mathbf{1}_{\{y(t)=y'\}} + \theta^{y(t)} + \theta^{m(t)} + \theta^{r} + \mathbf{X}_{it} \cdot \gamma + v_{it} \tag{2}$$

where $H_i = 1$ indicates that the physician is operating in a high viral infection prone region ("High Viral Region") and $\theta^{y(t)}$, $\theta^{m(t)}$, and θ^{r} capture yearly, monthly and regional variation that is common to all physicians. $\mathbf{X}_{it}$ includes patient and disease profile variations across physicians and time as well as physician age and gender. I construct the incoming rate of diseases by the group based on the justifiability of antibiotic

[25]The reason I do not define the outcome at the week level is simply to have enough observations for each physician to have a more robust measure of performance.

[26]An alternative specification would take emergency room visits for bronchitis as a continuous treatment or "dose" and interact the variable with before and after 2017. However, recent literature on difference-in-differences with continuous treatment shows that the two-way fixed effect estimator associated with this specification is difficult to interpret Callaway et al. (2021). Nonetheless, for interested readers, the results of the estimation where I take the mean weekly ER visit rates as the continuous treatment is provided in Appendix A

treatment and I interact these variables with year fixed effects so that changes across years around the treatment of these diseases are accounted for. I also control for patient profiles using information on the distribution of age and different sets of chronic diseases. Additionally, I control for month-year fixed effects to capture differences in aggregate disease profiles across time and for region of operation of the physician, together with the age and gender of the patient.

The identification of the effect requires a parallel trends assumption together with homogeneity assumptions and no selection based on unobservables[27]. Since there are multiple pre-periods to change in bonus design, we can test the parallel trends assumption by testing $\beta_{2015} = \beta_{2016} = 0$. Under the null hypothesis, there is no difference between the treatment and control groups, absent the change in the design. Table 4 presents the estimates for Eqn. (2). The first and second columns use the rate of antibiotic prescriptions in levels. Since the dependent variable is the ratio of the number of antibiotic treatments to the number of patients treated, column 2 reports the results where observations are weighted by the number of patients. Finally, column 3 uses the logit transformation of the dependent variable to account for the nonlinearity of proportion measures.

[27]One important concern is the number of patients treated in a given month. Note that the dependent variable is the prescription rate, that is the number of antibiotic treatments per patient treated. If there is a change in the number of patients treated, this could inflate or deflate the dependent variable regardless of the change in prescription behavior. I rule this out by repeating the estimation for the number of patients treated. The results are presented in Section A.1.

Table 4: Estimation results of Eqn. (2)

	(1)	(2)	(3)
High viral region	0.0265***	0.0302***	0.169***
	(0.00147)	(0.00155)	(0.0105)
High viral region × Year 2015	-0.00322	-0.00326	-0.0238
	(0.00177)	(0.00192)	(0.0125)
High viral region × Year 2016	-0.00140	-0.00336	-0.0126
	(0.00176)	(0.00191)	(0.0121)
High viral region × Year 2017	-0.00505**	-0.00718***	-0.0290*
	(0.00172)	(0.00187)	(0.0121)
High viral region × Year 2018	-0.00667***	-0.00931***	-0.0399**
	(0.00172)	(0.00186)	(0.0122)
High viral region × Year 2019	-0.00551**	-0.00773***	-0.0259*
	(0.00172)	(0.00185)	(0.0125)
Mean dep. variable	0.191	0.191	.
Observations	80003	80002	80003

Notes: (1) Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
(2) The first column reports the results from the linear probability model, the second column from the weighted linear probability model, and the third column from the logit model.
(3) Coefficients for physician age and gender, Year, Month, Region FE's, rate of *sometimes* class diseases and *always* class diseases with Year FE interactions, 10th, 25th, 45th, 50th, 55th, 75th and 90th percentiles of patient age, chronic disease counts, and the share of eligible patients are not reported.

In all the model specifications, it is consistently observed that physicians practicing in regions with a higher prevalence of viral infections tend to have higher prescription rates, on average. However, the introduction of the P4P program in 2017, which required improved performance for the same level of reward, brought statistically significant changes. Physicians operating in regions with high viral infection rates decreased their prescription rates by an additional 0.5 to 0.72 percentage points compared to their counterparts in low viral infection-prone regions. To put this into perspective, the mean of the dependent variable is 18.8%, meaning that this additional decrease translates to an average reduction of 3.8%. While this effect might appear relatively modest, it is important to consider potential heterogeneity in physicians' responses, as discussed in Markovitz and Ryan (2017). To explore this potential heterogeneity, I examine two dimensions: physician age groups and their initial performance. I use 2014 values to classify physicians into three groups for each dimension, based on the 25th and 75th percentiles of the variable's distribution[28].

Table 5 presents the heterogeneity of the effect across physician age groups using the weighted linear probability model as the preferred specification. The findings highlight significant heterogeneity. Physicians

[28]For physician age, the 25th percentile is 47 and the 75th percentile is 58. For the initial performance, i.e., antibiotic prescription rate, the 25th percentile is 33% and the 75th percentile is 55%.

younger than 48 years are the best performers and also the most responsive, as presented in column 1. They not only start with the smallest difference in prescription rates based on their region of practice and the smallest average prescription rate overall, but they also exhibit a strong response to the incentives, almost offsetting the initial 1.8 percentage point difference in antibiotic prescription rates. This result is unsurprising, as their education and training likely align more closely with the growing concerns of antimicrobial resistance compared to older cohorts. Interestingly, physicians aged 48 to 58, although closer to the older cohort in their initial practices, also respond positively to the program, indicating that age alone does not necessarily dictate the response to incentives. Conversely, for the oldest physicians, the scenario is different. On average, older physicians practicing in regions with a higher viral infection rate tend to prescribe antibiotics at a higher rate, and there is no significant response to the change in the reward scheme. One possible explanation could be that older physicians treat older or sicker patients over time, although it is essential to note that the program and the measure used in this analysis specifically focus on patients younger than 65 years with no serious chronic diseases. Furthermore, a comprehensive set of controls approximates patient characteristics, which helps alleviate concerns about patient selection. Notably, the observed difference in prescription rates does not diminish from 2017 onwards, suggesting that the response to incentives, on average, is not significant. This finding aligns with existing literature on the cohort effect among physicians and their inappropriate use of antibiotics Cadieux et al. (2007); Mandelli et al. (2023).

Table 5: Estimation results of Eqn. (2) by physician cohort

	Physician Age		
	Less than 48	Between 48 and 58	More than 58
High viral region	0.0186***	0.0316***	0.0357***
	(0.00264)	(0.00225)	(0.00348)
High viral region × Year 2015	-0.00602	-0.00458	0.00297
	(0.00329)	(0.00279)	(0.00426)
High viral region × Year 2016	-0.00586	-0.00646*	0.00463
	(0.00324)	(0.00277)	(0.00431)
High viral region × Year 2017	-0.0102**	-0.0102***	0.00232
	(0.00316)	(0.00271)	(0.00421)
High viral region × Year 2018	-0.0160***	-0.0119***	0.00146
	(0.00304)	(0.00273)	(0.00414)
High viral region × Year 2019	-0.0135***	-0.0112***	0.00471
	(0.00306)	(0.00271)	(0.00417)
Mean dep. variable	0.178	0.190	0.207
Observations	20166	39803	20033

Notes: (1) Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
(2) Results are obtained from the weighted linear probability model.
(3) Coefficients for physician age and gender, Year, Month, Region FE's, rate of *Sometimes* class diseases and *Always* class diseases with Year FE interactions, 10th, 25th, 45th, 50th, 55th, 75th and 90th percentiles of patient age, chronic disease counts, and the share of eligible patients are not reported.
(4) Columns differ by the physicians in the sample based on their age in 2014. The cut-offs are the 25th and 75th percentiles of the distribution.

I conducted a similar analysis by examining the effect on physicians with high-, middle-, and low-performance levels in 2014. The results align with expectations. First, the baseline effect of practicing in a high-viral region becomes more pronounced as performance decreases, as seen from columns 1 to 3. The impact of the program is greatest for physicians with the lowest initial performance levels. However, the magnitude of the correction remains relatively small, with a 1 percentage point reduction in prescription rates compared to the baseline rate of 25.9%. For physicians with mid-range performance levels, there was already a decrease in prescription rates even before the bonus requirements became more stringent. This could be attributed to the fact that these physicians had the potential to improve their performance even under the rules from 2012 to 2016. Therefore, it is not possible to argue that parallel trends hold for this group. However, this does not necessarily negate the program's overall success. Overall, the effect of the program appears to be most pronounced for physicians with lower initial performance levels, while the impact on those with mid-range performance is less straightforward due to pre-existing trends.

Table 6: Estimation results of Eqn. (2) by physician performance in 2014

	Overall antibiotic prescription rate in 2014		
	Less than 33%	Between 33% and 55%	More than 55%
High viral region	0.0102***	0.0209***	0.0364***
	(0.00183)	(0.00153)	(0.00363)
High viral region × Year 2015	-0.00162	-0.00570**	0.00146
	(0.00227)	(0.00189)	(0.00336)
High viral region × Year 2016	-0.00137	-0.00405*	-0.000541
	(0.00231)	(0.00193)	(0.00333)
High viral region × Year 2017	-0.00150	-0.00608**	-0.00870**
	(0.00235)	(0.00189)	(0.00328)
High viral region × Year 2018	-0.00422	-0.00761***	-0.0143***
	(0.00231)	(0.00191)	(0.00325)
High viral region × Year 2019	-0.00560*	-0.00572**	-0.0101**
	(0.00220)	(0.00195)	(0.00329)
Mean dep. variable	0.126	0.190	0.259
Observations	19851	40156	19995

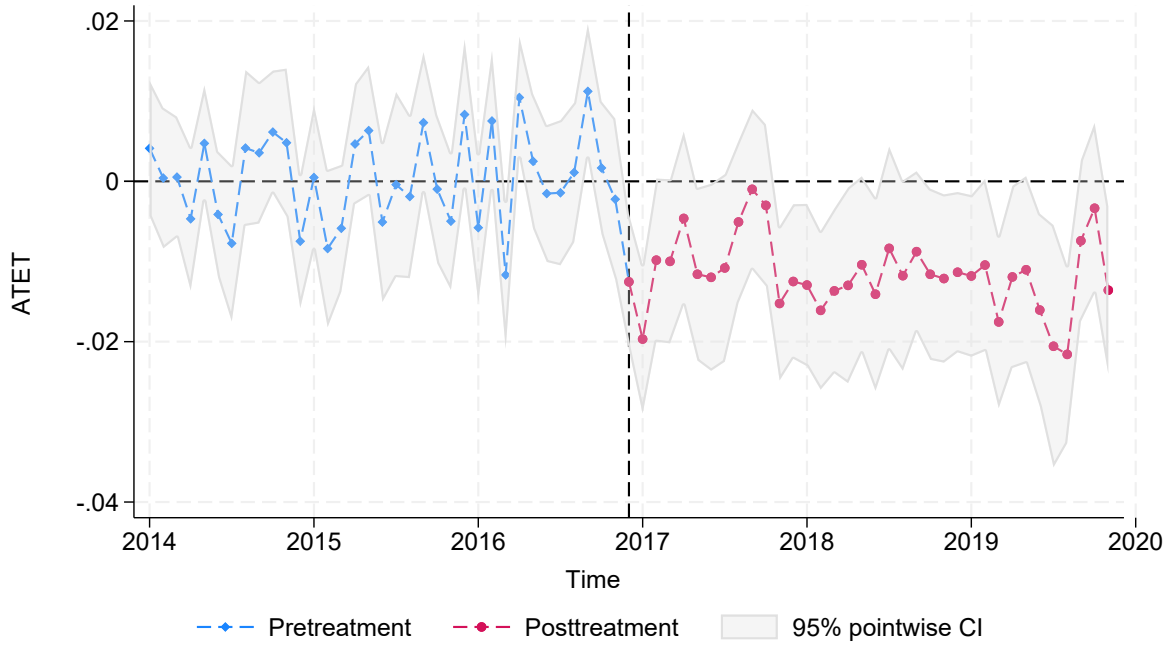
Notes: (1) Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
(2) Results are obtained from the weighted linear probability model.
(3) Coefficients for physician age and gender, Year, Month, Region FE's, rate of *Sometimes* class diseases and *Always* class diseases with Year FE interactions, 10th, 25th, 45th, 50th, 55th, 75th and 90th percentiles of patient age, chronic disease counts, and the share of eligible patients are not reported.
(4) Columns differ by the physicians in the sample based on their performance in 2014. The cut-offs are the 25th and 75th percentiles of the distribution.

The results from the standard linear regression point out heterogeneous treatment effects that depend on the physician-specific covariates such as physician age as presented in Table 5. This challenges whether the average treatment effect on the treated presented in Table 4 is unbiased. I address this issue in the next section.

3.1.1 Addressing the heterogeneous treatment effects

I re-estimate Eqn. 2 using regression adjustment as proposed by Heckman et al. (1997); Roth et al. (2023) to allow for covariate-specific trends. While there is no heterogeneity in the timing of treatment since the first period is January 2017 for all physicians, to capture the dynamic treatment effect at a more granular level I also allow the treatment effect to change across months. Average treatment effect on the treated at the monthly level are presented in Figure 11.

Figure 11: Dynamic ATET for the OAC

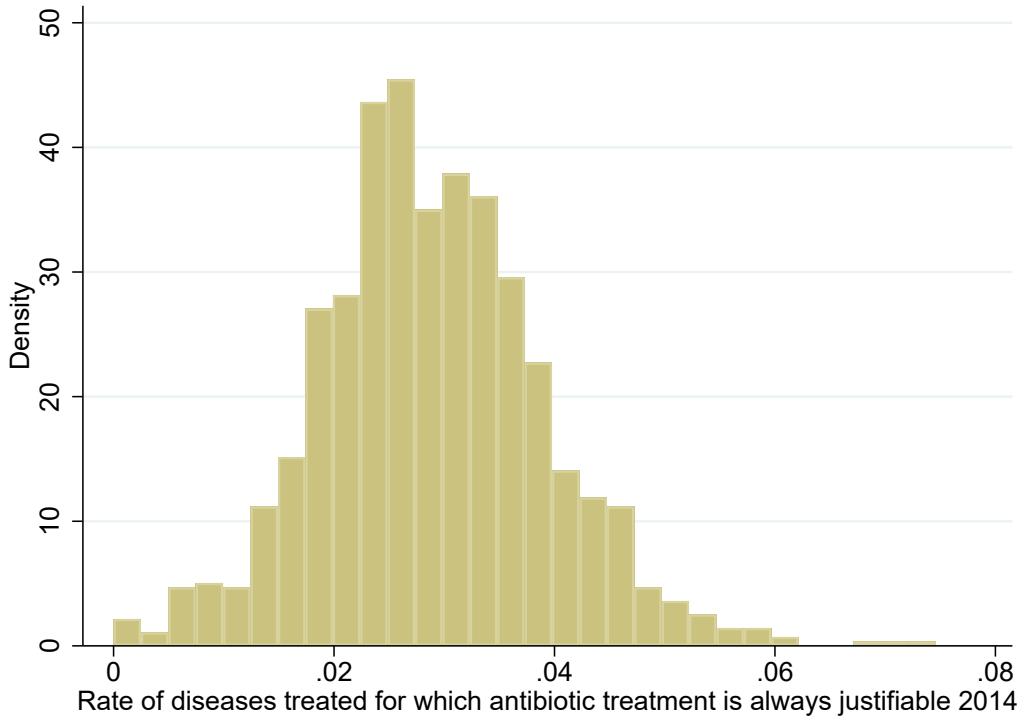


Notes: (1) Clustered standard errors at the physician level are presented.
(2) Coefficients for physician age and gender, Year, Month, Region FE's, rate of *sometimes* class diseases and *always* class diseases with time FE interactions, 10th, 25th, 45th, 50th, 55th, 75th and 90th percentiles of patient age, chronic disease counts, and the share of eligible patients are not reported.

3.2 At-risk antibiotics criterion

In this section, I use the variation in the disease profiles physicians treat to assign physicians into two groups: those who frequently treat diseases that almost invariably necessitate antibiotics and those who do so less frequently. The rationale behind this categorization is illustrated in Figure 3. Given the high rate of antibiotic prescriptions and the substantial proportion of at-risk antibiotics, it is reasonable to anticipate that physicians would respond more to the AAC criterion introduced in 2017 if they prescribed antibiotics more frequently for diseases within the *always* group. Figure 12 presents the distribution of incoming *always* group diseases across physicians in 2014.

Figure 12: Distribution of *always* class of diseases (%) across physicians in 2014



Notes: The figure presents the share of *always* diseases a physician encounters among visits in 2017. For example, 5% means that among all appointments that a physician has, 5% are related to diagnoses for which antibiotic prescription is for an *always* disease.

As presented in Figure 12, there is a large variation across physicians. Using this measure, I group physicians into two categories: those who face bacterial cases more frequently and less frequently. Physicians in the frequent bacterial cases group typically encounter diseases that almost *always* necessitate antibiotic treatment; therefore, they are also likely to have a larger share of "at-risk" antibiotic prescriptions, on average. If the physicians are responding to the introduction of the AAC, it would be more likely to be the physicians who provide treatment for more bacterial cases that would respond. This is what I test in the following analysis.

The outcome of interest is the share of at-risk antibiotics among all antibiotics prescribed by physician i at time t, denoted by Y_{it}. To construct the "pseudo" treatment and control groups, I use the median of the distribution of incoming *always* group diseases, which is 2.82%. The physicians whose frequency of patients with diseases in the *always* category, i.e., bacterial infections, is above this median are in the pseudo treatment group ($F_i = 1$, "Frequent Bacterial") and those with a frequency lower than the median are in the control group ($F_i = 0$). Above this level, the physicians are assigned to the frequent bacterial cases group,

denotes by the indicator F_i. The estimating equation is as follows.

$$Y_{it} = \alpha\ F_i + \sum_{y' \in \{2015,\ldots,2019\}} \beta^{y'} \cdot F_i \cdot \mathbf{1}_{\{y(t)=y'\}} + \theta^{y(t)} + \theta^{m(t)} + \theta^r + \mathbf{X}_{it} \cdot \gamma + v_{it} \tag{3}$$

where $\theta^{y(t)}$, $\theta^{m(t)}$, and θ^r capture yearly, monthly and regional variation common to all physicians and $\mathbf{X}_{it}$ includes patient and disease profile variations across physicians and time, as well as physician age and gender. Table 7 presents the estimates for Eqn. 3. The first column uses the rate in levels. Column 2 reports the results where observations are weighted by the number of antibiotic prescriptions. Finally, the third column uses the logit transformation of the dependent variable. In interpreting the results, note that the ratio of at-risk antibiotic prescriptions is also affected by the total number of antibiotics prescribed. In the previous section, I presented evidence that the overall prescription rates were affected by the change in the program in 2017. I argue that the effect of this simultaneous change makes the results stronger. If the physicians were not to change their at-risk antibiotic prescription rates and only decrease the prescriptions of not at-risk antibiotics, the AAC measure would naturally be inflated by a decrease in the denominator.

Table 7: Estimation results of Eqn. 3

	(1)	(2)	(3)
Frequent Bacterial	0.0357***	0.0364***	0.165***
	(0.00313)	(0.00355)	(0.0263)
Frequent Bacterial ×Year=2015	0.00153	0.00301	0.0757*
	(0.00423)	(0.00488)	(0.0352)
Frequent Bacterial ×Year=2016	-0.00194	0.00164	0.0210
	(0.00421)	(0.00488)	(0.0366)
Frequent Bacterial ×Year=2017	-0.0186***	-0.0169***	-0.0509
	(0.00414)	(0.00478)	(0.0385)
Frequent Bacterial ×Year=2018	-0.0183***	-0.0179***	-0.00806
	(0.00409)	(0.00460)	(0.0401)
Frequent Bacterial ×Year=2019	-0.0256***	-0.0215***	-0.0930*
	(0.00403)	(0.00454)	(0.0430)
Mean dep. variable	0.315	0.315	.
Observations	80114	80114	80003

Notes: (1) Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
(2) The first column reports the results from the linear probability model, the second column from the weighted linear probability model and the third column from the logit model.
(3) Coefficients for physician age and gender, Year, Month, Region FE's, 10th, 25th, 45th, 50th, 55th, 75th and 90th percentiles of patient age, chronic disease counts, and the share of eligible patients are not reported.

The results indicates a decrease in prescriptions of at-risk antibiotics after the introduction of the criterion

in 2017 for physicians who treat bacterial diseases at a higher rate. Therefore, the cuts in the at-risk antibiotics were significant enough to overcome the potential inflation of the measure due to the responses to the OAC criterion. As for the OAC criterion, using the multiple pre-periods, we can rule out the existence of a parallel trend before the introduction of the criterion. The effect is consistent across specifications while the statistical significance is weaker for the logit specification.

I repeat this exercise for different age cohorts, as defined previously. The results are presented in 8. The significance of the results across age groups follow the same patterns as for the OAC. The effect is more substantial for the younger physician cohorts.

Table 8: Estimation results of Eqn. (3) by age cohorts

	Less than 48	Between 48 and 58	More than 58
Frequent Bacterial	0.0293***	0.0472***	0.00344
	(0.00642)	(0.00510)	(0.00696)
Frequent Bacterial ×Year=2015	-0.00290	0.00493	0.00981
	(0.00876)	(0.00702)	(0.00949)
Frequent Bacterial ×Year=2016	-0.00922	0.00304	0.0102
	(0.00890)	(0.00699)	(0.00938)
Frequent Bacterial ×Year=2017	-0.0240**	-0.0190**	-0.000283
	(0.00851)	(0.00694)	(0.00917)
Frequent Bacterial ×Year=2018	-0.0426***	-0.0126	0.00510
	(0.00822)	(0.00657)	(0.00904)
Frequent Bacterial ×Year=2019	-0.0264***	-0.0209**	-0.00352
	(0.00796)	(0.00652)	(0.00907)
Mean dep. variable	0.319	0.316	0.310
Observations	20193	39856	20065

Notes: (1) Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
(2) Results are obtained from the weighted linear probability model.
(3) Coefficients for physician age and gender, Year, Month, Region FE's, 10th, 25th, 45th, 50th, 55th, 75th and 90th percentiles of patient age, chronic disease counts, and the share of eligible patients are not reported.
(4) Columns differ by the physicians in the sample based on their age in 2014. The cut-offs are the 25th and 75th percentiles of the distribution.

The same exercise is repeated for different groups based on initial performance[29]. The decrease is the most pronounced for the group that is in the mid-range performance. Contrary to the findings for the OAC presented in Table 9, the poor performers, i.e. physicians with 2014 performance is 49% or higher, do not present a significant change in their behaviors.

[29] As for the OAC, I took 25th and 75th percentile as thresholds.

Table 9: Estimation results of Eqn. (3) by initial performance

	Less than 30%	Between 30% and 49%	More than 49%
Frequent Bacterial	0.00616	0.0140***	0.00518
	(0.00383)	(0.00305)	(0.00598)
Frequent Bacterial ×Year=2015	0.00272	-0.000398	0.0122
	(0.00523)	(0.00431)	(0.00831)
Frequent Bacterial ×Year=2016	0.00129	0.00412	0.0121
	(0.00560)	(0.00440)	(0.00873)
Frequent Bacterial ×Year=2017	-0.0141*	-0.0183***	0.00618
	(0.00591)	(0.00449)	(0.00869)
Frequent Bacterial ×Year=2018	-0.0106	-0.0213***	0.0129
	(0.00546)	(0.00452)	(0.00836)
Frequent Bacterial ×Year=2019	-0.0133*	-0.0225***	0.0127
	(0.00554)	(0.00448)	(0.00853)
Mean dep. variable	0.206	0.304	0.439
Observations	19765	39655	20694

Notes: (1) Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
(2) Results use weighted linear probability model.
(3) Coefficients for physician age and gender, Year, Month, Region FE's, 10th, 25th, 45th, 50th, 55th, 75th and 90th percentiles of patient age, chronic disease counts, and the share of eligible patients are not reported.
(4) Columns differ by the physicians in the sample based on their performance in 2014. The cut-offs are the 25th and 75th percentiles of the distribution.

The findings for both criteria suggest that the incentive program has a significant effect in the expected areas. The OAC is linked to reductions in unnecessary antibiotic prescriptions, and the AAC is more pronounced among physicians who primarily prescribe antibiotics for bacterial diseases. However, it is challenging to precisely quantify the overall effect due to the absence of a clearly defined control and treatment group in a randomized setting. Furthermore, while there is evidence of heterogeneity, influenced by physician cohort and prior performance, a unified framework is required to gain a deeper understanding of how these incentives influence physician behavior. In the next section, I propose a simple model of antibiotic prescriptions to better understand how physicians respond to the bonus.

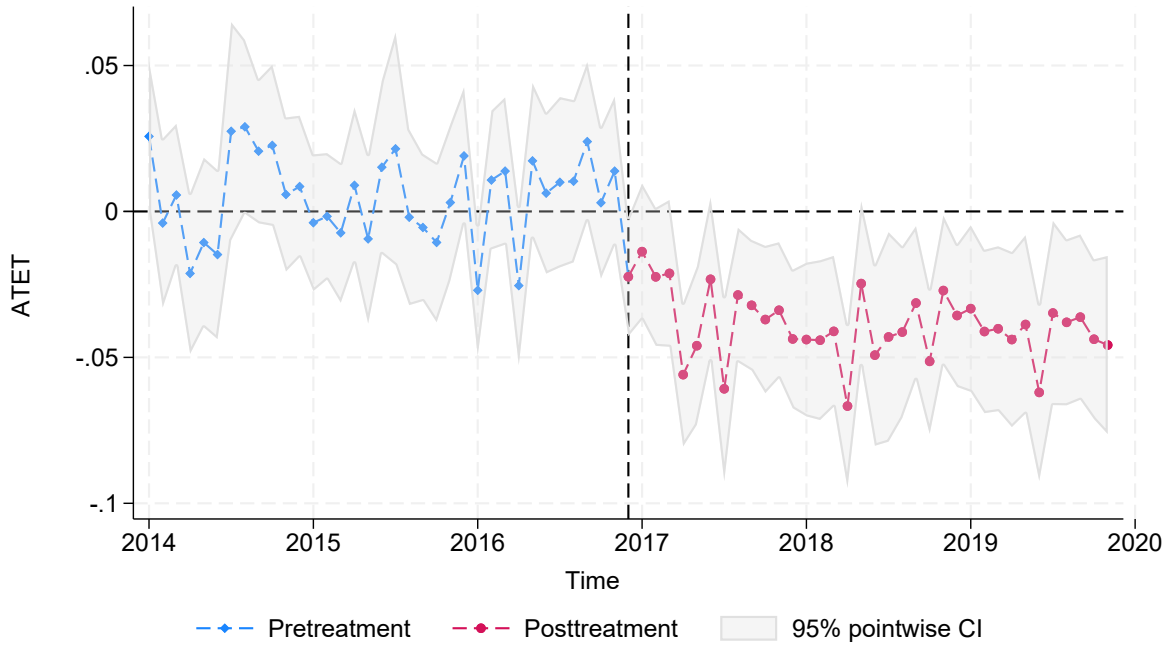
The results from the standard linear regression point out heterogeneous treatment effects as on the case for the OAC. I address this issue in the next section.

3.2.1 Addressing the heterogeneous treatment effects

I re-estimate Eqn. 3 using regression adjustment as proposed by Heckman et al. (1997); Roth et al. (2023) to allow for covariate-specific trends. As in the OAC case, there is no heterogeneity in the timing of treatment since the introduction of the criterion is January 2017 for all physicians. Average treatment effect on the

treated at the monthly level are presented in Figure 13.

Figure 13: Dynamic ATET for the AAC



Notes: (1) Clustered standard errors at the physician level are presented.
(2) Coefficients for physician age and gender, Year, Month, Region FE's, rate of *sometimes* class diseases for low-risk patients with time FE interactions, 10th, 25th, 45th, 50th, 55th, 75th and 90th percentiles of patient age, chronic disease counts, and the share of eligible patients are not reported.

4 Model of antibiotic prescriptions with pay-for-performance

In this section, I construct a model to analyze antibiotic prescriptions, considering the influence of the incentives generated by the P4P program. I consider the decision of antibiotic treatment, abstracting away from the type of antibiotics prescribed. In this respect, the focus is on the OAC criterion. The model sheds light on two critical aspects of physician behavior. First, it predicts that physicians are likely to approach the treatment of various diseases differently, depending on the type and frequency of these diseases. Second, the model highlights the significance of the bonus scheme's design, especially when physicians adopt a forward-looking perspective.

4.1 Baseline model

In this framework, each physician, denoted by i, makes prescription decisions over a predefined time period, and their reward is contingent on the total number of prescriptions at the end of this period, which is denoted

as T. During each period, a patient arrives for treatment. Each patient, indexed by j, is characterized by the type of disease, denoted by k, based on whether antibiotic treatment is justified and the severity of the disease. There is a benefit attached to the antibiotic treatment to the patient, which the physician internalizes through an altruism term. In each period, a physician must decide on the rules for prescribing antibiotic therapy to a patient, taking into account the severity, denoted as s_j, and the disease type, k. In the final period, the physician is rewarded a bonus $B(n)$, where n is the number of antibiotics prescribed. In the following, I detail the setting and the decision process.

Patient utility In each period, patient j arrives with a disease $k \in \{(\text{A})\ always, (\text{S})\ sometimes, (\text{N})\ never\}$ and a severity health state (severity of the disease) $s_j \in [0,1] \sim F_s$. The incoming disease of type k at time t arrives with λ_t^k. I allow for serial correlation by allowing the type to evolve following a Markov process $\mathbb{P}(k'|k) \geq 0.5$ for $k = k'$. Without loss of generality, I assume F_s follows a uniform distribution between 0 and 1. Upon receiving antibiotic therapy, the patient's utility is given by $h(s_j, k)$, which is increasing in s_j for all disease types, $\frac{\partial h(s_j,k)}{\partial s_j} > 0\ \forall k$. I normalize the utility from a non-antibiotic treatment to zero for all k such that $h()$ captures the relative benefit[30]. The utility from an antibiotic treatment depends on the severity and type of the disease in the following ways. First, for any severity except $s_j = 0$, an antibiotic prescription is the most valuable for diseases in the *always* group, followed by *sometimes* and *never*: $h(s_j, A) > h(s_j, S) > h(s_j, N)\ \ \forall s_j \in (0,1]$, where $h(1,A) > h(1,S) > h(1,N) > 0$. For $s_j = 0$, i.e., the patient is not sick with any disease, the antibiotic prescription is harmful due to side effects. This implies $h(0,A) = h(0,S) = h(0,N) < 0$. Second, the marginal value of antibiotic prescription increases in the same order, $\frac{\partial h(s_j,A)}{\partial s_j} > \frac{\partial h(s_j,S)}{\partial s_j} > \frac{\partial h(s_j,N)}{\partial s_j}\ \ \forall s_j \in [0,1]$. Finally, for each k $\exists s_k^*$, such that it solves $h(s_k^*, k) = 0$. Note that for each draw of severity above this threshold, it is optimal for the patient to receive an antibiotic treatment. The threshold that solves the optimal decision problem for the patient follows $s_N^* > s_S^* > s_A^*$ by the ordering of h(.,k).

Physician utility The physicians' utility consists of (i) the private cost/benefit attached to antibiotic prescription, (ii) the patient utility, and (iii) the financial incentives. A physician has two alternatives for the treatment decision, $a \in \{0\ (\text{No antibiotic treatment}),\ 1\ (\text{Antibiotic treatment})\}$[31]. I define the value of

[30]The function h() captures a combination of real health benefits and perceived ones by the patient. Antibiotic prescriptions in a situation where the cause of the infection is viral could be rationalized by a strong demand from the patient or an anticipation of a secondary infection that would be bacterial. I consider all the scenarios driven by patient utility to be captured by h(s) through s, and the physician derives utility from making the patient healthy/happy. Another rationalization would be that the physician observes a noisy signal of s, which allows him/her to assign an upper bound, $\bar{s}$. Learning the true value is costly, and if the physician does not want to incur the cost, s/he treats the patient as if the realized s is the upper bound $\bar{s}$. Then, the physician treats the patient based on the maximum value of s. Suppose the physician thinks the patient has sinusitis but does not know if the cause is bacterial or viral. A prescription could be rationalized by the physician treating the patient as if he or she is sure that it is bacterial. Finally, it could be rationalized by the physician liking the antibiotic regardless of the realized s, which can be due to habits, incorrect beliefs, etc.

[31]One can extend the model by using three alternatives: $a \in \{0$ (No antibiotic treatment), 1 (Antibiotic treatment with other than at-risk antibiotics), 2(Antibiotic treatment with at-risk antibiotics)$\}$. The

antibiotic treatment relative to the lack of antibiotic treatment. Therefore, absent the incentives, the utility from antibiotic treatment net of no antibiotic treatment for treating a patient with disease k and severity s_j is given by:

$$U(s_j, k) = b + \alpha \cdot h(s_j, k)$$

where α represents the degree of altruism of the physician, $\alpha \cdot h(s_j, k)$ captures the part of patient j's utility for the treatment of disease k that the physician internalizes, and b captures the antibiotic prescription being the "easier" option due to several reasons that include patient satisfaction and pressure, time constraints (limited time to see patients, diagnose their illnesses, and formulate a treatment plan, avoid lengthy explanations of why the drugs are not needed), and decision fatigue.

Bonus The incentives are assumed to be piece wise-linear such that if physician i prescribes antibiotic x at time T, the bonus $B(x)$ is given simply by

$$B(x) = \begin{cases} 1 & \text{if } 0 < x < \underline{N} \\ \frac{\overline{N} - x}{\overline{N} - \underline{N}} & \text{if } \underline{N} < x < \overline{N} \\ 0 & \text{if } x > \overline{N} \end{cases} \tag{4}$$

where $\overline{N}$ and $\underline{N}$ are exogenously set by the policy. This introduces dynamics to the physician's problem, where the decision is based on observable states n: the number of prescriptions realized and t the current period.

Treatment choice The flow utility of the physician at time t for treating disease k upon introduction of the financial incentives is as follows:

$$\begin{aligned} U^f(s_t^*|n, t, k) &= \mathbb{E}[\max\{U(s_t^*, k), 0\}] + \mathbb{1}_{\{t=T\}}\beta \ \mathbb{E}[B|n, t] \\ &= \mathbb{1}_{\{t=T\}}\beta \int_0^{s_t^*} \mathbb{E}[B|n, t+1] f(s) ds + \int_{s_t^*}^1 (b + \alpha \ h(s, k) + \mathbb{1}_{t=T}\beta \, \mathbb{E}[B|n+1, t+1]) \, f(s) ds \end{aligned}$$

assumptions would be modified to have at-risk antibiotics be more costly (stronger side effects) when severity is 0 but have higher value when severity is 1.

We can then write the lifetime discounted (with discount factor δ) utility at time t as follows [32]:

$$\begin{aligned} v(n,t,k) &= max_{s'_{n,t,k}} \int_{s'_{n,t,k}}^{1} (b+\alpha h(s,k)) f(s) ds \\ &\quad +\delta \sum_{k'} \mathbb{P}(k'|k) \left(F(s'_{n,t,k}) v(n,t+1,k') + (1-F(s'_{n,t,k})) v(n+1,t+1,k') \right) \end{aligned}$$

The optimal threshold solves the following equation:

$$h(s^*_{n,t,k},k) = \frac{-b+\delta \cdot \sum_{k'} \mathbb{P}(k'|k) \overbrace{(v(n,t+1,k') - v(n+1,t+1,k'))}^{\geq 0}}{\alpha} \tag{5}$$

Absence of financial incentives In the absence of financial incentives, the decision-making process is not influenced by intertemporal concerns, as there is no difference between the value of treatment in subsequent periods: $v(n+1,t+1,k') - v(n,t+1,k') = 0$ in Eqn. (5). Therefore, when patients arrive each period from the same distribution, the same threshold applies. This threshold is determined by the condition where the value function equals a certain negative threshold, specifically $h(s,k) = \frac{-b}{\alpha}$.

As the private benefits (b) of prescribing antibiotics increase, patients with less severe conditions are more likely to receive antibiotic prescriptions. However, this is discounted by the altruism of the physician, denoted as α. In the case of an infinitely altruistic physician, the decision to prescribe antibiotics aligns with maximizing the patient's utility, which only happens when the benefit of treatment exceeds the cost, meaning $h(s,k) = 0$. Furthermore, the assumption about the function $h(.)$ ensures that when antibiotic treatment is for an *always* disease, prescriptions are more frequent. This is followed by diseases in the *sometimes* group and then the *never* group. This aligns with the inherent justifiability of antibiotic treatment for different disease types.

4.2 How do financial incentives affect prescription behavior?

The financial incentives shift the optimal threshold of severity above which physicians prescribe antibiotics such that it satisfies Eqn. (5). This equality guides us in terms of how the prescription behavior should change under the set of assumptions of the model.

Physicians can be either forward-looking or myopic in their decision-making. Forward-looking physicians consider the future and form expectations about their future performance and the "price" or marginal bonus lost due to antibiotic prescription. In this context, they act based on their expectations, which can

[32] See Appendix B for the derivation.

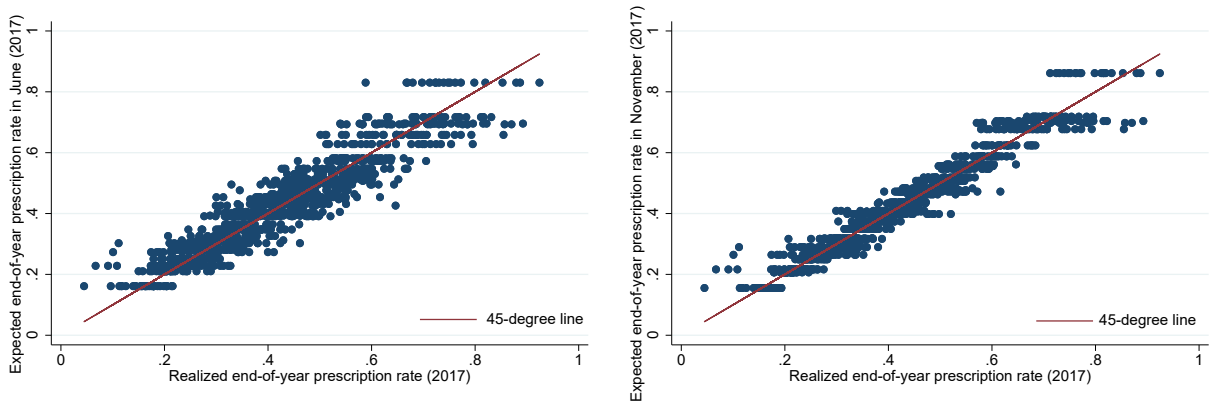
be represented as the difference between the expected bonuses at the beginning and end of the period: $E[B(.)|n, t+1] - E[B(.)|n+1, t+1]$. However, due to the inherent randomness associated with patient visits and the causes of these visits, there is a different level of uncertainty between the expectations at the beginning and end of the period. On the other hand, myopic physicians focus solely on the cost of prescription at the time of making the prescription, even though the realization of bonuses occurs at the end of the year. In the upcoming analysis, I explore how physicians react to bonuses, considering both forward-looking and myopic perspectives, and examine the impact of specific design features on their decision-making.

4.2.1 Expected end-of-year performance

If the physicians are forward-looking, the effect of bonuses on prescriptions goes through the marginal cost of prescription at the end-of-year performance. To capture the role of incentives, I construct the expected end-of-year performance in a way similar to Brot-Goldberg et al. (2017) who study consumer responses to medical care prices under a nonlinear insurance scheme. To group physicians with similar characteristics, I divide them into categories based on their "core" attributes, which remain constant over time, and their prescription patterns up to a particular month, denoted as month m.

For the core characteristics, I consider factors such as the physician's prescription rate, the number of patients eligible for the OAC, the share of *always* visits, and the share of *sometimes* visits, all of which are assessed based on data from the year 2014. I employ k-means clustering to classify physicians into 6 distinct groups. The use of 2014 data ensures that the core attributes reflect as little as possible the influence of the P4P program. Within each (core) group of physicians and for each month m, I further categorize physicians into 6 groups based on their prescription rates up to month m to capture the variation in realized prescriptions. For instance, two physicians with different pools of diagnoses might reach month m with different prescription rates already realized even though they are in the same core group. Since the prescriptions accumulate over a year period, this measure affects the expectations over what is achievable by the end of the year. As a result, at any given time, there are 36 distinct physician groups. Figures 14a and 14b illustrate the relationship between the expected end-of-year performance and the realized performance in June and November 2017, respectively.

Figure 14: Expected end-of-year performance versus realized end-of-year performance in 2017



(a) June (b) November

Note: Each point in the graph represents a physician in the given month, and the red line marks the 45 degree line.

Within each group, I calculate the expected end-of-year bonus using the average realized end-of-year prescription rate. Additionally, I consider the slope of the bonus function at that particular prescription rate. Notably, this slope effectively represents the cost to the physician associated with antibiotic treatments or the gains achieved from refraining from antibiotic prescriptions. Therefore, the slope captures $\mathbb{E}[B|n+1, t+1] - \mathbb{E}[B|n, t+1]$ for a forward-looking physician who factors in the threshold through differences in the value functions in Eqn. (5).

4.2.2 Bonus design and kink points

Decisions in each period are affected by the other periods' decisions through the mere existence of the bonus at the end period. Physicians take into account their anticipated performance at the end of the year, and their current actions influence the bonus they will receive. The effect is captured by the second term in Eqn. (5), which is what generates the deviation from the threshold in the absence of financial incentives. If physicians' prescription decisions are affected by the shape of the bonus function, the decisions should be smooth functions of expected performance, except the crossing points of kink points.

To see this, consider the last period, where the threshold of severity for antibiotic prescription is given by

$$h(s^*_{n,T,k}, k) = \frac{-b + \beta(B(n) - B(n+1))}{\alpha}.$$

The way the incentives are designed, there are three values of marginal change in bonus $(B(n) - B(n+1))$ depending on n. Either it is a constant bonus region ($n < \underline{N}$ or $n > \overline{N}$) and hence the bonus does not have an effect on prescriptions, or the state of the physician in terms of prescription falls into the constant

slope regions. Then, we expect that the bonus shifts the threshold to a higher value, thereby decreasing the prescription rate if β is nonzero and $B(n) - B(n+1) > 0$. Note that $s^*_{n,t,k}$ then has two values:

$$s^*_{n,T,k} = \begin{cases} h^{-1}\left(\frac{-b}{\alpha}, k\right) & \text{if } n < \underline{N} \text{ or } n > \overline{N} \\ h^{-1}\left(\frac{-b+\beta(\frac{1}{\overline{N}-\underline{N}})}{\alpha}, k\right) & \text{if } n > \underline{N} \text{ and } n < \overline{N} \end{cases} \tag{6}$$

The points of discontinuity for $s^*_{n,T,k}$ are located at $n = \underline{N}$ and $n = \overline{N}$. These points are similar to the target and intermediary thresholds in the P4P program, as they create kink points in the design. If physicians are forward-looking and they perfectly predict the future, they might not react differently when they cross the threshold. For example, if a physician anticipates that they will have a zero bonus slope throughout the year, there may be no incentive for them to alter their behavior before or after the threshold. However, the random nature of diseases and visits makes the expected end-of-year performance a random object. This randomness is depicted in Figures 14a and 14b, where we observe that predictions become more precise as the year progresses, from June to November. Therefore, if physicians are indeed learning from crossing the threshold, we would expect changes in their prescription behavior in response to this new information. In other words, the shift in the expected end-of-year slope is associated with crossing the threshold. During the sample period, these kink points change twice, in 2017 and then in 2018 (see Table 1).

4.2.3 Differential effects to different disease groups under financial incentives

The threshold rule outlined in Equation 5 demonstrates how physicians balance patient utility $h()$ (on the left-hand side) with their own private benefits b and bonuses (on the right-hand side). The variation in the two components drives the change in response to the bonus program. This leads to two predictions that can be empirically tested in terms of the varying impacts of the bonus across different diseases.

The first prediction comes from the heterogeneity across diseases and the value of antibiotics in their treatment $(h(s, \text{always}), h(s, \text{sometimes}), h(s, \text{never}))$. The assumptions on the ordering of these functions together with the assumption on the first derivatives suggests that incentives affect various diseases in distinct ways. In particular, a unit change on the right-hand side of Equation 5 is expected to have the most substantial impact on diseases where antibiotic treatment is for a *never* class disease since the marginal value of antibiotic treatment is the lowest for this group. This is followed by diseases categorized as *sometimes* and those classified as *always*. However, an unintended consequence of the program is that if physicians place too much importance on financial incentives, the program could lead to a decrease in prescriptions when they are necessary. Using the prescription-level data, I estimate the disease-specific impact of the financial incentives in Section 5 to test these predictions.

The second prediction pertains to the significance of diseases, which is closely linked to their prevalence and how it impacts the private benefits associated with antibiotic prescriptions for physicians. Up to this point, the model assumed uniform private benefits. In this section, I propose a way in which private benefits can vary across diseases, affecting how physicians respond to bonuses. Let us consider two diseases within the same category, labeled as "k" – Disease 1 and Disease 2. The probability of a patient arriving with Disease 1 is denoted as p_1, while the probability for Disease 2 is denoted as p_2, where it holds that $p_1 > p_2$. These probabilities are independent of past occurrences. The key distinguishing factor between these diseases is the private benefits associated with antibiotic prescriptions. In certain cases, physicians might face constraints such as the need to explain treatments to patients or decision fatigue. Under these circumstances, they may find it more convenient to respond to incentives for diseases that occur more frequently. This is because doing so reduces the overall time and energy required for prescriptions, especially in the case of Disease 1. Another explanation is that the salience of a disease makes it more prominent in the decision-making process, while other diseases are given less weight. While I may not be able to distinguish definitively between these explanations, I aim to test whether the salience of a disease plays a role in influencing physician behavior.

In Section 5.1, I use this variation to show that physicians responds to the decrease in marginal returns around the kink point.

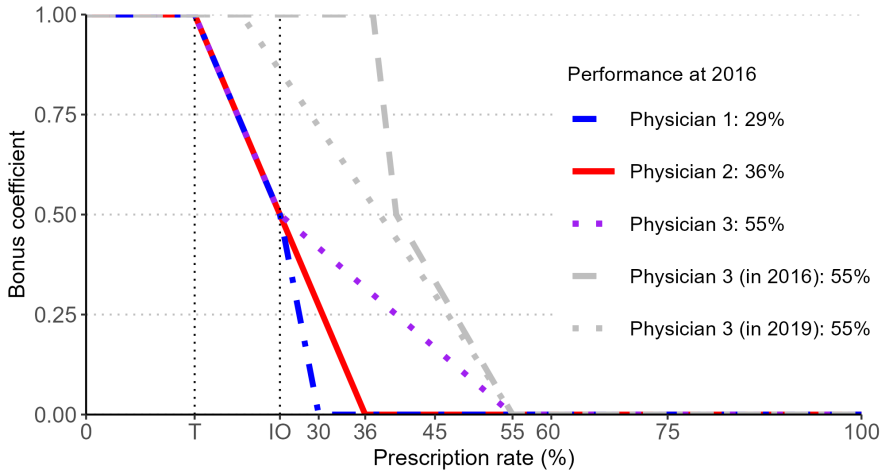
5 Empirical analysis and results

5.1 Forward-looking behavior and the impact of kink points in bonus design

In 2017, the bonus qualification requirements were made more stringent. To receive the full bonus, doctors were mandated to lower their antibiotic prescription rate from 37% to 14%. The intermediary objective set by the program was also reduced. In 2018, the bonus program underwent further changes, maintaining a stricter target rate while introducing a more lenient intermediary objective. I use these changes in the criteria to study the impact of bonuses. These changes led to two variations that I exploit in this section. First, the slope of the bonus function that physicians face changed, generating variation in the marginal cost of prescriptions. Second, the intermediary threshold in 2017 was placed in a way that for other years, the slope of the bonus function is constant. Therefore, this enables testing of whether physicians respond to such a kink point created by the policy, as we can use other years to control for the possible confounding factors that could occur at the time of the year where the physician reaches the score associated with a drop in the marginal bonus reduction (i.e., marginal cost of prescription) of an antibiotic prescription in 2017. To present these changes graphically, Figure 15 plots the bonus coefficient in 2017 for three physicians that differ in their prescription rates in 2016. Physician 1 prescribed to 29% of the (low-risk) patients in 2016,

whereas this rate was 55% for physician 3. Moreover, to see the changes in the bonus design, physician 3's bonus coefficient functions in 2016 and 2019 are also provided.

Figure 15: Bonus design across years in the sample period

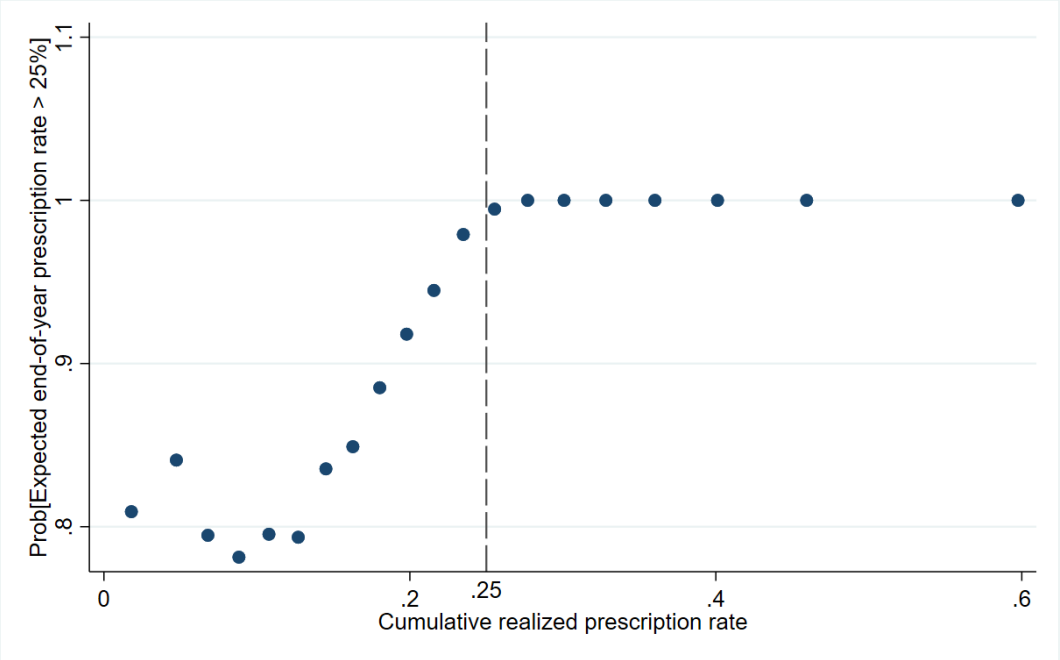


Notes: (1) The colored lines plot the bonus coefficient in 2017 for three physicians who have different reference year performance. Physician 1's reference year performance is 29%, Physician 2's is 36% and Physician 3's is 55%. To provide the bonus coefficient before 2017, I plot the bonus function for Physician 3 in year 2016 (dashed gray line). To provide the bonus coefficient after 2017, I plot the functions for Physician 3 in year 2019 (dotted gray line). The goal is to visually present the absence of a kink point at 25%, which is the intermediary objective in 2017 in other years.
(2) T denotes the target rate in 2017, IO denotes the intermediary objective in 2017.

Two observations are important for understanding the exogenous variation created by the changes in the bonus design. First, for each physician in the sample, the bonus function changes generate different incentives for (non-)antibiotic prescription. This is important as in a given year, the bonus itself is endogenous to performance. Second, there is a kink point at 25% in the bonus function for 2017, while for 2018-19, the bonus at 25% is downward sloping, and in 2014-16, it is zero without any kink point. This implies that upon crossing 25%, physicians have lower incentives if their performance was greater than 36% (which is the rate that ensures a constant slope at the kink point) in 2016. Absent the change in slope, i.e., marginal bonus reduction upon antibiotic prescription, we expect to have no effect in the other years upon crossing 25%. I leverage these two properties of changes in the design to examine whether physicians react to the expected marginal cost of prescription and to determine if the kink point has a significant effect on behavior. To better grasp physicians' predictions about exceeding 25% of their year-end prescriptions (with details on

how these predictions are calculated in Section 4.2.1) and the informativeness of exceeding 25%, I conduct the following exercise. Using the expected end-of-year performance of physicians, I plot the proportion of physicians who have an expected end-of-year prescription rate higher than 25% within bins of the realized cumulative prescription rate in Figure 16. I use data from 2014 to 2016 in order to not include the effects of the changes in the bonus design. In line with the existing prescription patterns, even for low values of realized prescription rate, the expectations of physicians to pass the threshold is quite high. The proportion increases to 100 percent upon passing the threshold in a steep manner, indicating that approaching and passing the threshold conveys information about the future. Hence, reacting to the design specifics goes together with the forward-looking behavior given the uncertainty around prescriptions and the realizations throughout the year.

Figure 16: Expected performance across the realized prescription rates

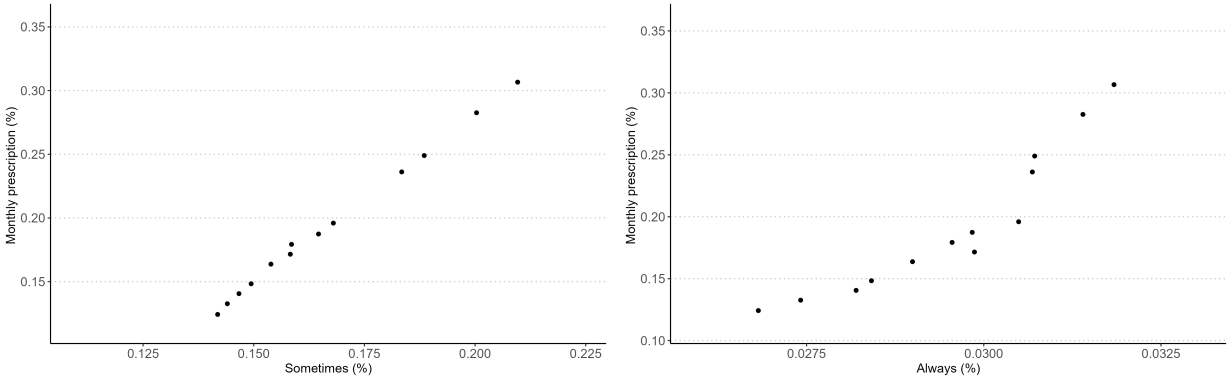


Note: The x-axis marks the number of antibiotic prescriptions already realized as a percentage of the patients in the given year. The y-axis presents the unconditional averages of the expected end-of-year performance being larger than 25% across physicians in the corresponding bin in terms of the cumulative prescription rates.

Note that crossing of the 25% threshold is influenced by the overall prescription behavior of the physician, making it an endogenous variable. Physicians who place a higher value on antibiotic treatment are more likely to cross the 25% threshold, and they are also more inclined to prescribe more in the absence of specific incentives. Consequently, endogeneity introduces downward bias into the estimates of the effect at the kink point. Additionally, if physicians have high expectations about crossing this threshold throughout the year,

they may opt to prescribe at a higher rate even before reaching the threshold, in line with forward-looking behavior. While Figure 16 indicates that some uncertainty remains just before reaching the 25% threshold, using instruments helps to address the endogeneity problem by exploiting exogenous factors that make physicians more or less likely to pass the threshold. These instruments involve the history of incoming diseases, which vary in their justifiability for antibiotic treatment. I calculate the cumulative share of incoming diseases for which antibiotic treatment is for either an *always* or *sometimes* disease. The numerator represents the total number of incoming diseases within a specific group until month m in year y, while the denominator represents the total number of visits. Physicians with a higher number of *sometimes* or *always* cases are more likely to have higher antibiotic prescription rates compared to physicians treating cases where antibiotic treatment is for a *never* disease. Figures 17a and 17b illustrate the unconditional averages of prescription rates in relation to diseases in the *sometimes* and *always* groups across physicians and months, supporting the use of such instruments. The accumulation of these diseases influences the timing of crossing the 25% threshold earlier or later in the year based on the realized cases, but it does not independently affect the prescription behavior of physicians outside of the bonus structure.

Figure 17: Binned scatter plot of average monthly prescription rates across incoming disease rates (2014-2019)



(a) *Sometimes* group diseases (b) *Always* group diseases

In the following, Y_{it} represents the antibiotic prescription rate for physician i at time t. $Slope_{it}$ denotes the anticipated slope of the bonus function at the end of the year, which captures the expected loss in bonus upon prescription. Meanwhile, M_{it}^{25} defines the month of the year of time t at which physician i exceeds the 25% threshold. This indicator is instrumental in conveying information about end-of-year performance. The estimating equation is as follows.

$$\mathrm{Y}_{it} = \beta \cdot \mathrm{Slope}_{it} + \sum_{y' \in \{2014,\ldots,2019\}} \alpha_{y'} \cdot \mathbf{1}_{\{m(t) \geq M_{iy(t)}^{25}\}} \cdot \mathbf{1}_{y'=y(t)} + \underbrace{\boldsymbol{X}_{iy(t)}^{phy} \boldsymbol{\beta}^{phy}}_{\text{Physician Char.}} + \underbrace{\boldsymbol{X}_{it}^{pt} \boldsymbol{\beta}^{pt}}_{\text{Patient \& Case}} + \gamma_{m(t)} + \gamma_{y(t)} + \epsilon_{it} \quad (7)$$

To account for heterogeneity among physicians, I incorporate factors such as the gender and age of the

physician. Additionally, as before, I include month (γ_m) and year (γ_y) controls. To address differences in the time required to treat a patient and variations in income due to their practice, I consider the number of patients, especially high-risk ones, and the annual number of patient visits. Furthermore, I control for differences in patient profiles and the types of cases treated. I take into account the mean age of treated patients, the average number of chronic diseases, and the 10th, 25th, 45th, 50th, 55th, 75th and 90th percentiles of these distributions per physician and month. I also control for the monthly rates of cases that are *always* bacterial or sometimes bacterial. Finally, as shown in Figure 15, changes in incentives can influence the marginal returns of prescribing one less antibiotic. For physicians whose performance in the reference year falls below 36% but above 25%, their bonus decreases more rapidly upon crossing the threshold, making each saved antibiotic more valuable. In contrast, if their performance exceeds 36%, the incentives to prescribe fewer antibiotics are lower. Relatively few physicians are within this range, and I account for this by applying a fixed effect that identifies these specific physicians.

To isolate the effect, I interact the indicator of crossing 25% with a dummy variable that marks the 6-month interval around the crossing. The reason is that as time progresses, physician prescription behavior might be affected by other incentives related to the bonus design, which might introduce distortions around the crossing effect. I restrict the sample to physicians whose prescriptions exceed 25%, which includes approximately 85% of the physicians in the sample.

Table 10: Estimation results of Eqn. (7)

	OLS	IV1	IV2
E[Slope OAC]	0.000315***	0.000657***	0.000629***
	(0.0000773)	(0.0000941)	(0.0000924)
E[Slope OAC] × Year ≥ 2017	0.00309***	0.00236***	0.00233***
	(0.000338)	(0.000337)	(0.000332)
Passed 25% × Within 6 months	0.0371***	-0.0106	-0.00672
	(0.00151)	(0.00979)	(0.00949)
Passed 25% × Within 6 months × Year 2015	-0.00372	-0.0145*	-0.0132*
	(0.00214)	(0.00582)	(0.00518)
Passed 25% × Within 6 months × Year 2016	0.00174	0.00361	0.00293
	(0.00213)	(0.00549)	(0.00502)
Passed 25% × Within 6 months × Year 2017	0.00668**	0.0256***	0.0216***
	(0.00211)	(0.00562)	(0.00515)
Passed 25% × Within 6 months × Year 2018	-0.000802	0.00354	0.00285
	(0.00216)	(0.00560)	(0.00520)
Passed 25% × Within 6 months × Year 2019	0.00139	0.000840	0.000981
	(0.00214)	(0.00568)	(0.00526)
Within 6 months	0.0815***	0.0640***	0.0637***
	(0.00155)	(0.00209)	(0.00198)
Within 6 months × Year 2015	-0.00494*	0.00582	0.00509
	(0.00218)	(0.00330)	(0.00306)
Within 6 months × Year 2016	0.00113	-0.00112	-0.000779
	(0.00216)	(0.00309)	(0.00292)
Within 6 months × Year 2017	-0.00936***	-0.0159***	-0.0140***
	(0.00210)	(0.00309)	(0.00291)
Within 6 months × Year 2018	-0.00164	-0.00232	-0.00192
	(0.00214)	(0.00310)	(0.00295)
Within 6 months × Year 2019	-0.00946***	-0.00562	-0.00561
	(0.00214)	(0.00311)	(0.00295)
Observations	65694	65673	65673
p-value of Underidentification LM statistic		3.24e-135	4.67e-134
p-value of Hansen J statistic		0.0000219	0.000484

Notes: (1) Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
(2) IV1: Cum. incoming share of *sometimes* groups diseases interacted with year FE's and *always* group diseases without year interactions, IV2: Cum. incoming share of *always* group diseases and *sometimes* diseases interacted with year FE's.
(3) The average monthly prescription rate in the sample before 2017 was 22.2% with a standard deviation of 8.1%.

The results are presented in Table 10[33]. The first observation is that the higher the expected marginal

[33]I repeat the analysis by using weights to control for the dependent variable being a ratio of number of antibiotics to the number of patients. The results are presented in Table 15

loss in bonus upon antibiotic prescription in magnitude is, the lower the antibiotic prescription rate. Note that since the incentive structure does not allow for too much variation, even across physicians before 2017, I allow for the effect of the bonus to differ before and after 2017. As expected, the impact is much greater starting from 2017. To understand the magnitude of the effect, if we take a physician who expects to be in between the target and the intermediary threshold in 2019 with the $Slope = -2.8$, the impact of the program is -0.9 pp fewer antibiotic prescriptions using estimates from column 3. The effect is quite small in size. In the next section, I show that the effect is heterogeneous across diseases. The second observation is that, on average, in the 6 months around the crossing of 25%, physicians prescribe less in 2017, where 25% corresponds to the intermediary objective, compared to the other years, where the target or intermediary objective is not 25%. However, also in 2017, crossing the 25% threshold led to 1.85 pp more antibiotic prescriptions compared to any other years in the sample. The effect is larger than the drop around the 25% threshold, which generates excessive prescriptions after the threshold.

5.2 Impact of the financial incentives by disease group

Providing evidence on forward-looking behavior, I proceed by using the *Slope* variable to estimate the decision to prescribe antibiotics for each diagnosis separately. This enables me to uncover potential heterogeneity in response across diseases and to test whether the predictions of the model are correct. To do that, I use the prescription data because of its granularity. This enables me to capture the variation in each treatment decision based on case and patient specifics. In this way, I aim to approximate the health status or severity of the case and the variation in the diseases to the best extent possible. I use data from 2017 and 2019, as these are the years where the policy generated exogenous variation in the bonus function. Moreover, unlike 2018, when new rules were announced in the middle of the year, the physicians know exactly which thresholds they should respond to.

To that end, physician i's utility from providing treatment $a \in \{0, 1\}$ in case c is as follows:

$$U_{ic}^{a} = \mathbf{x}_{iac}\boldsymbol{\alpha}_i + \mathbf{z}_{ic}\boldsymbol{\delta}_a + \epsilon_{ic}^{a}$$

where

– $\mathbf{x}_{iac}$ are treatment-specific variables for which the effect is captured by a random coefficient, allowing differentiated taste across physicians: In this case, it represents the expected end-of-year slope for each physician .

– $\mathbf{z}_{ic}$ are case-specific variables, corresponding to the following decomposition:

$$\mathbf{z}_{ic} = [\mathbf{z}_c^{patient}, \mathbf{z}_i, \mathbf{z}_{\{Month(c),\ Region(c)\}}]$$

The latent utility from prescription of physician i to case c is as follows:

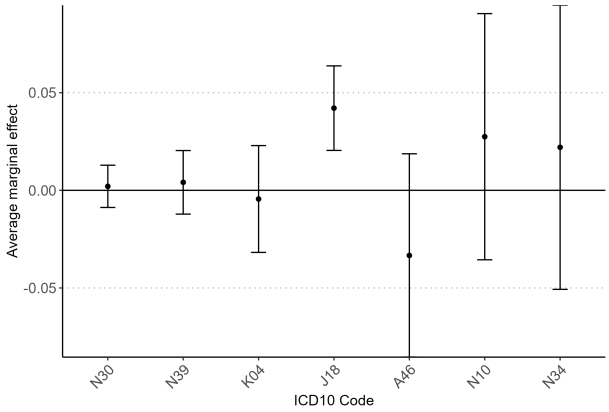
$$U_{ic}^1 = \beta Slope_{im(c)} + \underbrace{\boldsymbol{z}_c^{patient}}_{\text{Patient \& Case Char.}} \boldsymbol{\beta}^{patient} + \underbrace{\boldsymbol{z}_i^{phy}}_{\text{Physician Char.}} \boldsymbol{\beta}^{phy} + \underbrace{\gamma_{r(i),m(c)}}_{\text{Region x Month}} + \epsilon_{ic}^1 \tag{8}$$

where ϵ_{ic}^1 is assumed to follow the standard logistic distribution.

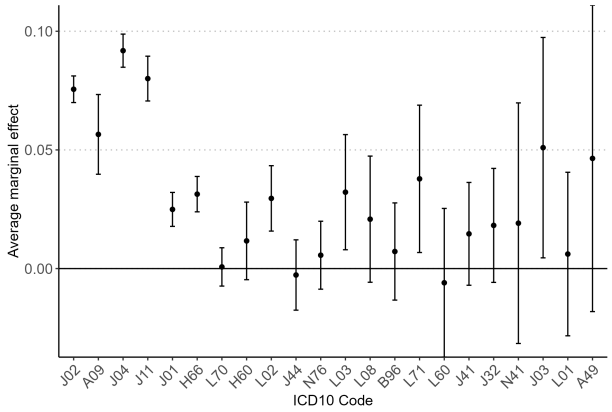
Assuming a common taste parameter, β, I estimate the logit model for each disease separately. In Figure 18, I present the average marginal effects of the *Slope* from the logit model for each disease, separated by disease groups. Figures 18a, 18b, and 18c present the results for the *always*, *sometimes* and *never* diseases, respectively. The results reveal heterogeneity in responses both across and within disease groups. First, in line with the predictions, physicians do not react significantly to the incentives for diseases that require antibiotic prescriptions, as seen in Figure 18a. One exception is pneumonia, where the category that specifies the cause of the disease is unknown. This indicates that the cause could be viral or bacterial, justifying the significant effect. The null result is a reassuring result with respect to concerns around the undesired effects of such incentives and signals that physicians are sufficiently altruistic. Furthermore, on average, the effect is larger for the *never* diseases than it is for the *sometimes*. Second, within each group for *sometimes* and *never*, I order the diseases by their frequency in the data and present the most frequent ones on the left side of the x-axis. On average, the more prevalent a disease is, the larger the impact of the policy. This could be due to a "salience effect", which is a cognitive bias in which individuals tend to focus their attention on the most prominent item. An alternative explanation would be economies of scale in the cost of change in the behavior. It could be that the cost of a behavior change decreases when implemented on a larger scale.

Figure 18: Average marginal effect of the expected slope from Eqn. (8) by disease groups

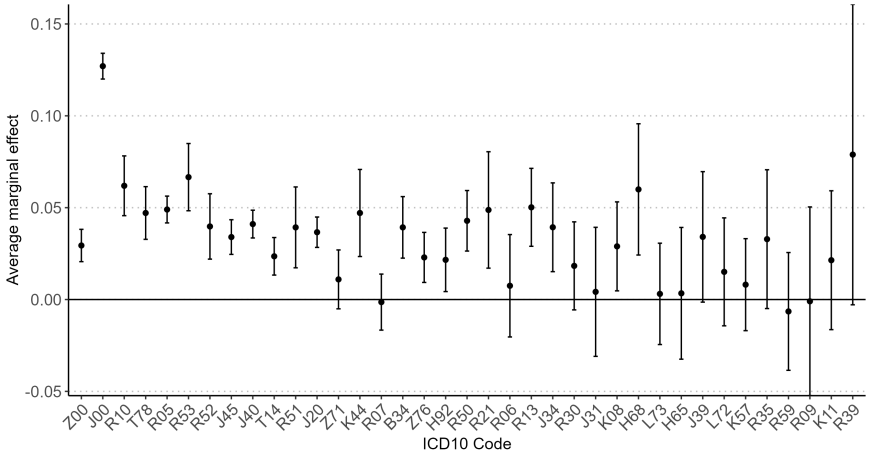
(a) *Always*



(b) *Sometimes*



(c) *Never*



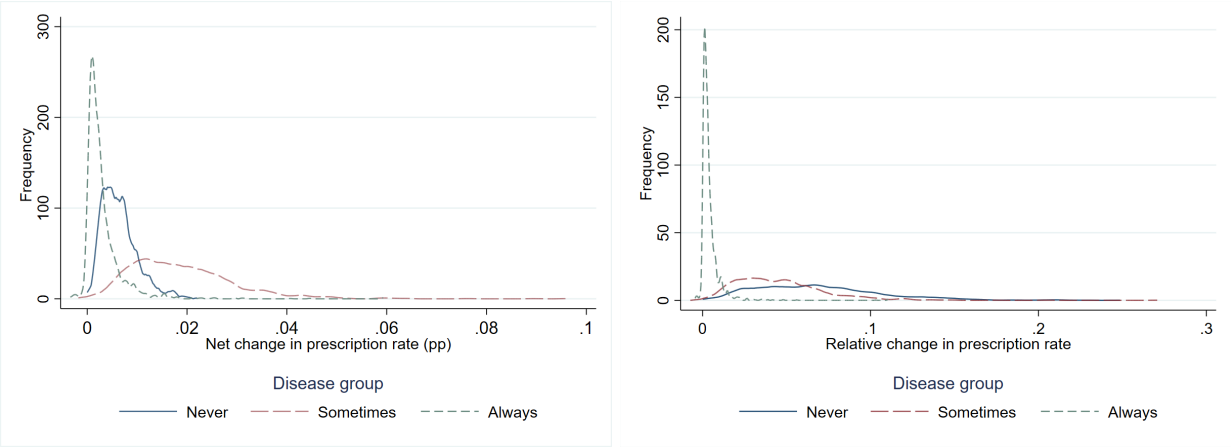
Notes: (1) Diseases within each disease group are ordered by their prevalence in the data in decreasing order from left to right.
(2) Top 4 diseases in *Always*: N30: Cystitis, N39: Other disorders of the urinary system, K04: Diseases of pulp and periapical tissues, J18: Pneumonia, unspecified organism
(3) Top 6 diseases in *Sometimes*: J02: Acute pharyngitis, A09: Other gastroenteritis and colitis of infectious and unspecified origin, J04: Acute laryngitis, J11: Influenza J01: Acute maxillary sinusitis, H66: Suppurative and unspecified otitis media
(4) Top 6 diseases in *Never*: Z00: General medical examination, J00: Acute nasopharyngitis (common cold), R10: Abdominal and pelvic pain, T78: Adverse effects, not elsewhere classified, R05: Cough, R53: Malaise and fatigue

5.3 Aggregate impact of the financial incentives and cost of the program

Using the estimates obtained in the previous section, I investigate the aggregate effect and the distribution of this effect. It is important for this purpose to consider that physicians treat a variety of diseases at different rates, and the impact per physician depends on their disease composition. I predict the prescription rates per disease with and without the bonus program for 2019. Then, I aggregate the effects of the diagnoses within

each disease group. The distribution of the net change and the relative change over physicians are presented in Figure 19. The net changes presented in Figure 19a show that in the aggregate, there is heterogeneity across physicians' responses within each group of diseases. The percentage point change is largest for the *sometimes* group, followed by the *never* group. Note however that the baseline prescription rates are different for these groups, as presented in Figure 3. Therefore, I also calculate the relative change as the ratio of the net change and the prescription rates absent the financial incentives, as predicted by the estimates. Figure 19b shows that, on average, the decrease in the *never* group is larger than that of the *sometimes* group. The change across physicians varies from 2% to almost 30%.

Figure 19: Distribution of net and relative changes in prescription rates across disease groups



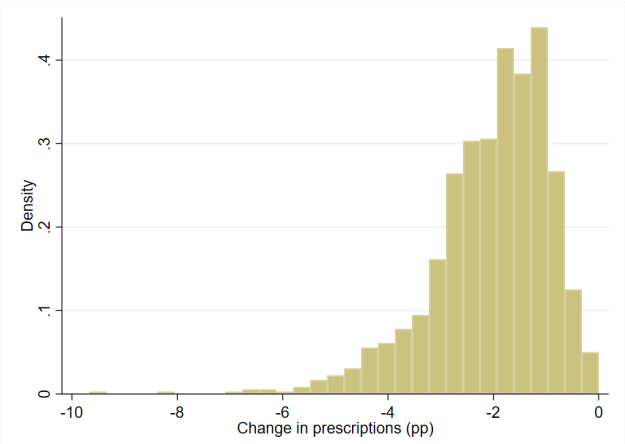
(a) Net change (pp)

(b) Relative change

Note: For each disease group, the distributions of the relative and net changes in prescriptions over physicians are provided for 2019. The net change is the difference between the prescription rates with and without the bonus program. The relative change is the ratio of the net change to the prescription rates absent the bonus.

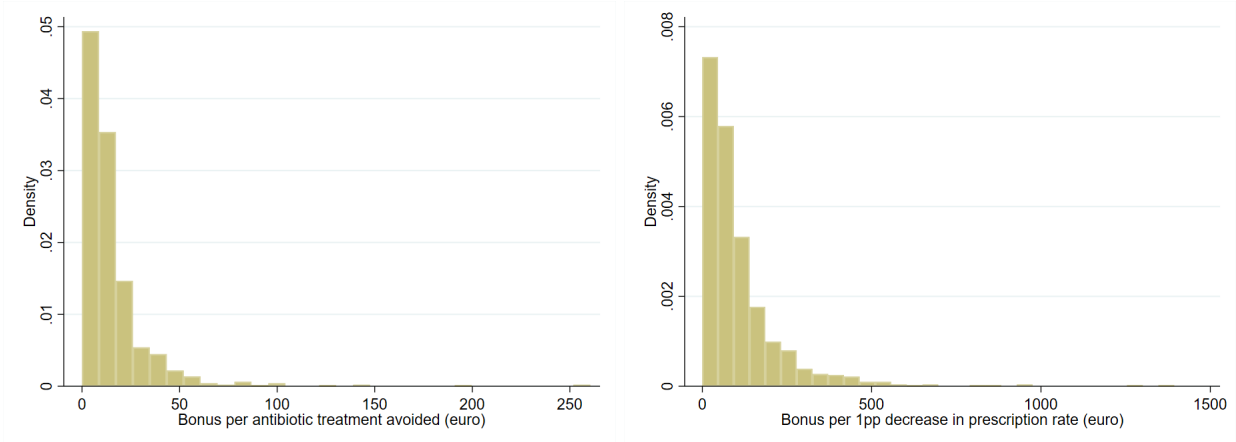
Figure 20 presents the resulting percentage point change in prescription rates across physicians in 2019.

Figure 20: Distribution of change in prescription rates (pp)



To determine the cost of these changes in prescriptions, I calculate two quantities that could be of interest to policymakers. The first is the bonus earned by the physician per (avoided) antibiotic prescription. The distribution across physicians is presented in Figure 21a. The distribution has a median of €10 and a mean of €14. Second, to understand how large or small these payments are, I also calculate the bonus paid for a one percentage point decrease in prescriptions, which is the quantity of interest of the program. It costs the health agency €106 per year, on average (median €73) to obtain a 1 percentage point decrease in prescriptions. Therefore, the average physician earned approximately €156 as a bonus payment with respect to the OAC criterion per year.

Figure 21: Distribution of bonuses

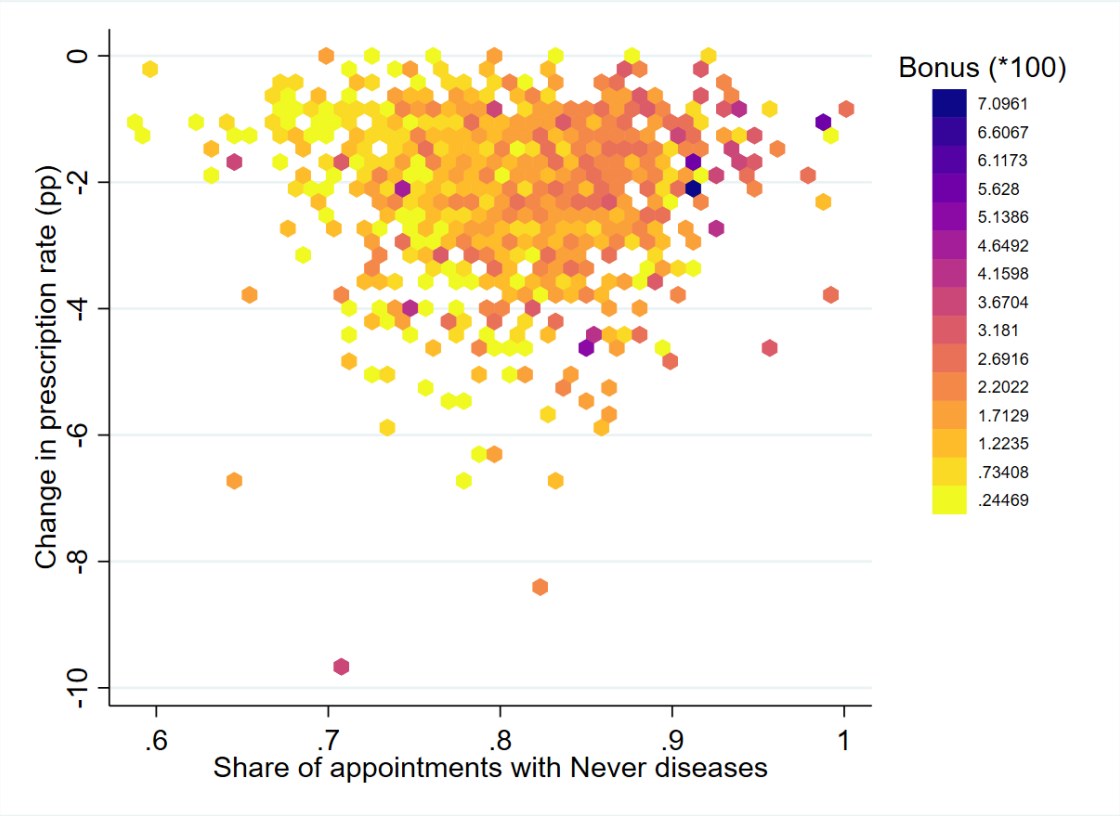


(a) Bonus rewarded (€) for an avoided antibiotic prescription in 2019

(b) Bonus rewarded for a percentage point decrease in 2019

Notably, the policy rewarded physicians based on their prescription rates, not their actual improvement in rates. Therefore, even if a physician did not improve, they would still earn a bonus based on their prescription rate. The fairness of the program might require that physicians who are already prescribing at low rates, i.e., those already performing better from the public health perspective, also earn bonuses. This would justify not rewarding improvement but rather prescription rates directly. However, the reason for low prescription rates is important because this policy might end up rewarding the mix of diseases treated by physicians, given the inherent variations across different disease groups, rather than focusing on actual rate improvements. As a result, this could create limitations in the impact of the policy for physicians who are performing moderately well. To obtain an idea of the extent to which this could be happening, I plot the bonus payments and the improvement across the share of visits of *never* groups of diseases.

Figure 22: Bonus across disease composition and change in prescriptions



Notes: The y-axis is the predicted change due to financial incentives, and the x-axis is the share of visits associated with a *never* group disease. The color maps the average bonus earned in the corresponding hexagon. The darker the color, the higher the bonus is.

Figure 22 illustrates that when considering a fixed "improvement" (i.e., a change in prescription rate), the bonus increases with the share of visits for diseases categorized as *never*. Furthermore, when examining the share of visits involving *never* class diseases, there is not a substantial difference in the bonuses earned across various rates of improvement. This figure emphasizes that bonus payments are determined by (at least) two key factors: the degree of improvement and the types of diseases that physicians encounter in their practice.

6 Conclusion

In this paper, I study the effectiveness of financial incentives within the pay-for-performance (P4P) program and provide a framework to understand where the impact occurs and how the design of the incentives matters. I aim to offer insights into how these incentives impact the prescription practices of general practitioners, with the overarching goal of protecting public health while efficiently managing public finances.

To this end, the results highlight the crucial role of well-designed financial incentives in influencing physician behavior regarding antibiotic prescriptions. Two primary performance criteria were established in the P4P program: the Overall Antibiotics Criterion (OAC) and the At-risk Antibiotics Criterion (AAC). These criteria target inappropriate antibiotic usage and aim to protect the efficacy of high-value antibiotics against resistant bacteria, respectively. Exploiting changes in the bonus function design over time, coupled with the varying prevalence of viral infections in different regions and the pools of diseases physicians treat, I show that physicians increase their performance more when they have room for improvement in response to more stringent requirements for performance pay. To better understand how this impact is generated, in the second part of the paper, I propose a model of treatment choice, taking into account the financial incentives introduced by the P4P program. This model revealed that the design of the bonus function matters, especially if physicians are forward-looking and respond to dynamic incentives. Moreover, the response to these incentives should vary depending on the necessity of antibiotic prescription for different diseases. In 2019, the financial incentives led to a reduction in prescriptions for diseases that do not necessitate antibiotics, resulting in an approximately 7% decrease. For diseases in the *Always* group, the reduction was approximately 0.3%. While the program's effect may be limited, this aligns with the associated costs. On average, physicians earned approximately €156. However, noting that the bonus is not conditional on improvement but on the realized prescription rates, the cost of the program per antibiotic prescription saved is not negligible. Each antibiotic prescription saved is compensated by €14. Notably, there is significant variation in these figures, primarily driven by the demographics of patients and the specific diseases they present.

The findings offer insights into how financial incentives play a role in shaping physician behavior in the context of antibiotic prescription practices. Thus far in the P4P program, the selected thresholds apply to everyone. Moreover, this selection appears to be guided by the budget constraints on the program, as for every criterion, the thresholds correspond to the same percentiles. The results highlight two dimensions of tailoring that could be of interest to policymakers. First, the pool of diseases physicians treat plays an important role in defining the magnitude of improvement that a physician can achieve, as well as their baseline bonus earnings. It is multidimensional in the sense that the diseases that lead to overprescription of antibiotics are not the same as the diseases that particularly cause the overprescription of at-risk antibiotics. Although it could be difficult in practice for policymakers to oversee the actual cause of prescriptions, measures such as the proneness to viral infections or the average composition of patients treated by physicians could be used as a tailoring tool. Second, physician characteristics, especially the cohort effect, matter for the success of the program. The results highlight the inertia from older physicians, whereas the younger cohorts are shown to be more responsive. However, the younger cohorts have better initial performance. Therefore, it could

be worth investigating the inertia among older physicians in their antibiotic prescriptions and eventually introduce either complementary attempts or explore other incentive mechanisms, such as payments upon improvement as opposed to the current system, where even absent an improvement, physicians still earn bonuses. Finally, this paper considers the decision to prescribe an antibiotic treatment and abstracts away from the choice of antibiotics. In Dubois and Gokkoca (2023), the focus is solely on the choice of antibiotics and the role of bacterial resistance for the treatment of cystitis (which almost *always* requires antibiotic treatment). Future research could combine the two approaches in a unified framework to create a more targeted approach regarding the effect of the policy for each antibiotic group.

References

Aron-Dine, A., L. Einav, A. Finkelstein, and M. Cullen (2015, October). Moral Hazard in Health Insurance: Do Dynamic Incentives Matter? *Review of Economics and Statistics 97*(4), 725–741.

Brot-Goldberg, Z. C., A. Chandra, B. R. Handel, and J. T. Kolstad (2017). What does a deductible do? the impact of cost-sharing on health care prices, quantities, and spending dynamics. *The Quarterly Journal of Economics 132*(3), 1261–1318.

Cadieux, G., R. Tamblyn, D. Dauphinee, and M. Libman (2007). Predictors of inappropriate antibiotic prescribing among primary care physicians. *Cmaj 177*(8), 877–883.

Callaway, B., A. Goodman-Bacon, and P. H. Sant'Anna (2021). Difference-in-differences with a continuous treatment. *arXiv preprint arXiv:2107.02637*.

Carlet, J., V. Jarlier, J. Acar, O. Debaere, P. Dehaumont, B. Grandbastien, P. Le Coz, G. Lina, Y. Pean, C. Rambaud, et al. (2020). Trends in antibiotic consumption and resistance in france over 20 years: Large and continuous efforts but contrasting results. In *Open Forum Infectious Diseases*, Volume 7.

Carlet, J. and P. Le Coz (2015). Together, Let's Save Antibiotics, proposals of the special working group for keeping antibiotics effective.

Carlet, J. and P. Le Coz (2015). Tous ensemble, sauvons les antibiotiques, Propositions du groupe de travail spécial pour la préservation des antibiotiques. Available at `https://www.vie-publique.fr/rapport/35184-tous-ensemble-sauvons-les-antibiotiques-propositions-du-groupe-de-tra` (2023/11/13), English: `http://solidarites-sante.gouv.fr/IMG/pdf/rapport_carlet_anglais.pdf` (2018/06/12).

Chevreul, K., K. Berg Brigham, I. Durand-Zalesk, and C. Hernández-Quevedo (2015). France: Health system review. *Health Systems in Transition 17*(3), 1–218.

Chua, K.-P., M. A. Fischer, and J. A. Linder (2019). Appropriateness of outpatient antibiotic prescribing among privately insured us patients: ICD-10-CM based cross sectional study. *BMJ 364*.

Chung, D. J., T. Steenburgh, and K. Sudhir (2014). Do bonuses enhance sales productivity? A dynamic structural analysis of bonus-based compensation plans. *Marketing Science 33*(2), 165–187.

Cole, A. (2014, August). GPs feel pressurised to prescribe unnecessary antibiotics, survey finds. *BMJ 349*, g5238. Publisher: British Medical Journal Publishing Group Section: News.

Constantinou, P., J. Sicsic, and C. Franc (2017). Effect of pay-for-performance on cervical cancer screening participation in france. *International journal of health economics and management 17*(2), 181–201.

Dalton, C. M., G. Gowrisankaran, and R. J. Town (2020). Salience, myopia, and complex dynamic incentives: Evidence from medicare part d. *The Review of Economic Studies 87*(2), 822–869.

DiMasi, J. A., L. Feldman, A. Seckler, and A. Wilson (2010). Trends in risks associated with new drug development: success rates for investigational drugs. *Clinical Pharmacology & Therapeutics 87*(3), 272–277.

Dowd, B., R. Feldman, and W. Nersesian (2013). Setting pay for performance targets: do poor performers give up? *Health Economics 22*(2), 168–179.

Dranove, D. (1988). Demand inducement and the physician/patient relationship. *Economic inquiry 26*(2), 281–298.

Dubois, P. and G. Gokkoca (2023). Antibiotic demand in the presence of antimicrobial resistance. *TSE Working Paper*.

Dubois, P. and T. Tunçel (2021). Identifying the effects of scientific information and recommendations on physicians' prescribing behavior. *Journal of health economics 78*, 102461.

Eijkenaar, F., M. Emmert, M. Scheppach, and O. Schöffski (2013). Effects of pay for performance in health care: A systematic review of systematic reviews. *Health policy 110*(2-3), 115–130.

Einav, L., A. Finkelstein, and P. Schrimpf (2015, May). The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D *. *The Quarterly Journal of Economics 130*(2), 841–899.

Ellegård, L. M., J. Dietrichson, and A. Anell (2018). Can pay-for-performance to primary care providers stimulate appropriate use of antibiotics? *Health Economics 27*(1), e39–e54.

Fleming-Dutra, K. E., A. L. Hersh, D. J. Shapiro, M. Bartoces, E. A. Enns, T. M. File, J. A. Finkelstein, J. S. Gerber, D. Y. Hyun, J. A. Linder, et al. (2016). Prevalence of inappropriate antibiotic prescriptions among us ambulatory care visits, 2010-2011. *Jama 315*(17), 1864–1873.

Fleming-Dutra, K. E., A. L. Hersh, D. J. Shapiro, M. Bartoces, E. A. Enns, T. M. File, Jr, J. A. Finkelstein, J. S. Gerber, D. Y. Hyun, J. A. Linder, R. Lynfield, D. J. Margolis, L. S. May, D. Merenstein, J. P. Metlay, J. G. Newland, J. F. Piccirillo, R. M. Roberts, G. V. Sanchez, K. J. Suda, A. Thomas, T. M. Woo, R. M. Zetts, and L. A. Hicks (2016, May). Prevalence of Inappropriate Antibiotic Prescriptions Among US Ambulatory Care Visits, 2010-2011. *JAMA 315*(17), 1864–1873.

Grubb, M. D. and M. Osborne (2015, January). Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock. *American Economic Review 105*(1), 234–271.

Guillemot, D., C. Carbon, F. Vauzelle-Kervroëdan, B. Balkau, P. Maison, G. Bouvenot, and E. Eschwège (1998, January). Inappropriateness and Variability of Antibiotic Prescription among French Office-Based Physicians. *Journal of Clinical Epidemiology 51*(1), 61–68.

Heckman, J. J., H. Ichimura, and P. E. Todd (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies 64*(4), 605–654.

HSA, U. (2022, November). English surveillance programme for antimicrobial utilisation and resistance (espaur) report 2021 to 2022.

Klein, T. J., M. Salm, and S. Upadhyay (2022). The response to dynamic incentives in insurance contracts with a deductible: Evidence from a differences-in-regression-discontinuities design. *Journal of Public Economics 210*, 104660.

Kohut, M. R., S. C. Keller, J. A. Linder, P. D. Tamma, S. E. Cosgrove, K. Speck, R. Ahn, P. Dullabh, M. A. Miller, and J. E. Szymczak (2020). The inconvincible patient: How clinicians perceive demand for antibiotics in the outpatient setting. *Family practice 37*(2), 276–282.

Lévin, C., N. Thilly, M. Dousak, G. Beraud, M. Klesnik, S. Uhan, D. Nathwani, B. Beovic, and C. Pulcini (2019, June). Perceptions, attitudes, and practices of French junior physicians regarding antibiotic use and resistance. *Médecine et Maladies Infectieuses 49*(4), 241–249.

Mandelli, S., I. Ardoino, A. Nobili, I. Fortino, and C. Franchi (2023). General practitioner-related factors associated with antibiotic prescription in community-dwelling adult population. *Pharmacoepidemiology 2*(2), 148–156.

Markovitz, A. A. and A. M. Ryan (2017). Pay-for-performance: Disappointing results or masked heterogeneity? *Medical Care Research and Review 74*(1), 3–78.

Misra, S. and H. S. Nair (2011). A structural model of sales-force compensation dynamics: Estimation and field implementation. *Quantitative Marketing and Economics 9*, 211–257.

Mullen, K. J., R. G. Frank, and M. B. Rosenthal (2010). Can you get what you pay for? Pay-for-performance and the quality of healthcare providers. *The Rand journal of economics 41*(1), 64–91.

Nevo, A., J. L. Turner, and J. W. Williams (2016). Usage-based pricing and demand for residential broadband. *Econometrica 84*(2), 411–443.

OECD (2023). Embracing a one health framework to fight antimicrobial resistance. *OECD Health Policy Studies*.

Or, Z., C. Gandré, A.-V. Seppänen, C. Hernández-Quevedo, E. Webb, M. Michel, and K. Chevreul (2023). Health system review. *Health 25*(3).

Richards, A. R. and J. A. Linder (2021). Behavioral economics and ambulatory antibiotic stewardship: A narrative review. *Clinical Therapeutics 43*(10), 1654–1667.

Rosenthal, M. B., R. G. Frank, Z. Li, and A. M. Epstein (2005). Early experience with pay-for-performance: From concept to practice. *Jama 294*(14), 1788–1793.

Roth, J., P. H. Sant'Anna, A. Bilinski, and J. Poe (2023). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*.

Sanchez, M.-A., S. Sanchez, L. Bouazzi, L. Peillard, A. Ohl-Hurtaud, and C. Quantin (2023). Does the implementation of pay-for-performance indicators improve the quality of healthcare? first results in france. *Frontiers in Public Health 11*, 1063806.

Schnell, M. (2017). Physician behavior in the presence of a secondary market: The case of prescription opioids. *Princeton University Department of Economics Working Paper*.

Sicsic, J. and C. Franc (2017). Impact assessment of a pay-for-performance program on breast cancer screening in france using micro data. *The European Journal of Health Economics 18*(5), 609–621.

Simon, M., N. Thilly, O. Pereira, and C. Pulcini (2022). Factors associated with the appropriateness of antibiotics prescribed in french general practice: A cross-sectional study using reimbursement databases. *Clinical Microbiology and Infection 28*(4), 609.e1–609.e6.

Sun, L., E. Y. Klein, and R. Laxminarayan (2012). Seasonality and temporal correlation between community antibiotic use and resistance in the united states. *Clinical infectious diseases 55*(5), 687–694.

Thilly, N., O. Pereira, J. Schouten, M. E. Hulscher, and C. Pulcini (2020). Proxy indicators to estimate appropriateness of antibiotic prescriptions by general practitioners: a proof-of-concept cross-sectional study based on reimbursement data, north-eastern france 2017. *Eurosurveillance 25*(27).

Van Herck, P., D. De Smedt, L. Annemans, R. Remmen, M. B. Rosenthal, and W. Sermeus (2010). Systematic review: effects, design choices, and context of pay-for-performance in health care. *BMC health services research 10*(1), 1–13.

Ventola, C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. *Pharmacy and therapeutics 40*(4).

WHO (2015). Global action plan on antimicrobial resistance. *Antimicrobial Resistance Division, National Action Plans and Monitoring and Evaluation, World Health Organization 2015*.

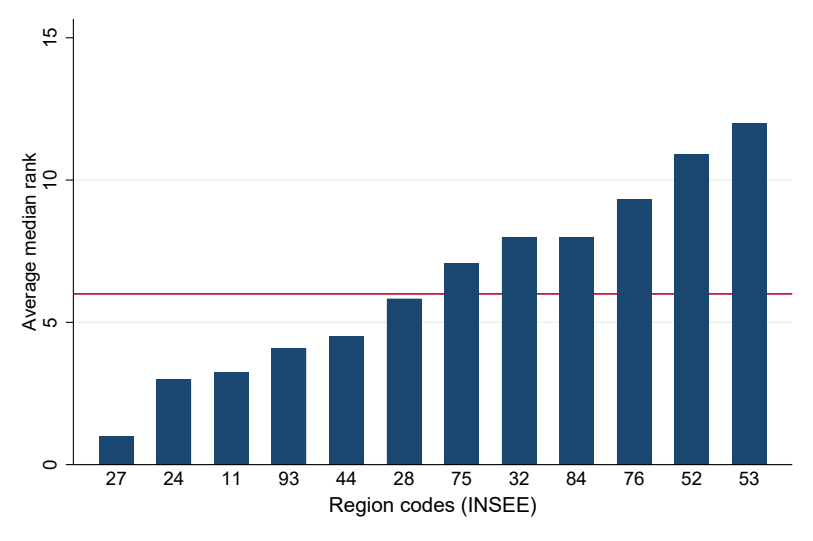
Yoshikawa, Y., I. Feldhaus, E. Özçelik, T. C. O. Hashiguchi, and M. Cecchini (2021). Financial strategies targeting healthcare providers to promote the prudent use of antibiotics: A systematic review of the evidence. *International Journal of Antimicrobial Agents 58*(6), 106446.

A Appendix

A.1 Additional results for OAC

Average median ranks across regions Average median ranks across regions are presented in Figure 23.

Figure 23: Average median ranks across regions



Continuous treatment In the following, instead of an indicator of high viral infection prone regions, I use a continuous treatment. For each month $m(t)$, I use the maximum weekly ER visit rate observed in the physician i's region of operation.

Table 11: Estimation results of Eqn. (2) with continuous treatment

	(1)	(2)	(3)
Max. ER rate	0.0148***	0.0192***	0.0883**
	(0.00428)	(0.00465)	(0.0282)
Year=2015 × Max. ER rate	-0.00799	-0.00912	-0.0523
	(0.00530)	(0.00572)	(0.0344)
Year=2016 × Max. ER rate	0.000289	-0.00661	0.00710
	(0.00575)	(0.00615)	(0.0373)
Year=2017 × Max. ER rate	-0.0135*	-0.0170**	-0.0835*
	(0.00550)	(0.00590)	(0.0359)
Year=2018 × Max. ER rate	-0.0141**	-0.0159**	-0.0823*
	(0.00547)	(0.00588)	(0.0368)
Year=2019 × Max. ER rate	-0.0129*	-0.0147*	-0.0709
	(0.00622)	(0.00682)	(0.0416)
Observations	80003	80002	80003

Notes: (1) Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
(2) First column reports the results from the linear probability model, second column from the weighted linear probability model and third column from the logit model.
(3) Coefficients for physician age and gender, Year, Month, Region FE's, rate of *Sometimes* class diseases and *Always* class diseases with Year FE interactions, 10th, 25th, 45th, 50th, 55th, 75th and 90th percentiles of patient age, chronic disease counts, the share of eligible patients are not reported.

Number of patients One confounding factor could be a change in number of patients which is the denominator in the calculation of the performance measure. If there are significant differences in the number of patients treated occurring at the same time as the change in incentives in 2017, this would lead to a change in the denominator and could inflate or deflate the observed performances regardless of the change in physician behavior. This would harm the interpretation of the findings as the effect of the pay-for-performance program. To understand if there is such confounding factor, I test whether there are any differences in patients treated before and after 2017 for treatment and control groups. I estimate Eqn. (2) where the dependent variable Y_{it} is the number of low-risk patients treated by physician i in month, $m(t)$. The results are presented in 12 for both the binary treatment and the continuous treatment cases. The results show that there is no significant change in the number of patients treated between the treatment and control groups at the time of the change in incentive design.

Table 12: Estimation results of Eqn. (2) with Y_{it} := Number of low-risk patients treated by physician i in month, m(t)

	Binary treatment	Continuous treatment
High Viral Region	10.61***	
	(1.282)	
High Viral Region × Year 2015	0.170	
	(1.527)	
High Viral Region × Year 2016	1.490	
	(1.540)	
High Viral Region × Year 2017	1.233	
	(1.532)	
High Viral Region × Year 2018	1.101	
	(1.528)	
High Viral Region × Year 2019	0.325	
	(1.530)	
Average ER rate		-5.311
		(3.810)
Average ER rate × Year 2015		4.659
		(4.538)
Average ER rate × Year 2016		4.982
		(5.019)
Average ER rate × Year 2017		4.253
		(4.634)
Average ER rate × Year 2018		-0.349
		(4.811)
Average ER rate × Year 2019		4.880
		(5.138)
Observations	80003	80003

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: (1) Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
(2) First column reports the results with binary treatment. Second column reports the results with continuous treatment.
(3) Coefficients for physician age and gender, Year, Month, Region FE's, rate of *Sometimes* class diseases and *Always* class diseases with Year FE interactions, 10th, 25th, 45th, 50th, 55th, 75th and 90th percentiles of patient age, chronic disease counts, the share of eligible patients are not reported.

A.2 Géodes data for Flu

I repeat the same exercise with ER visits due to flu rates instead of bronchitis. Notice that the flu is more concentrated on several weeks compared to the bronchitis.

Figure 24: Rate of visits to emergency room for flu

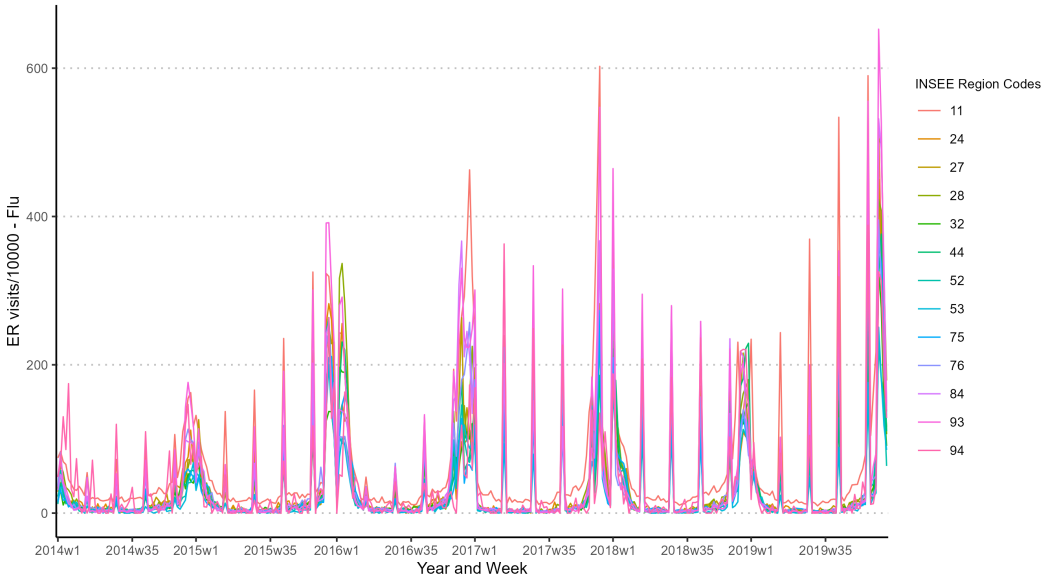


Figure 25

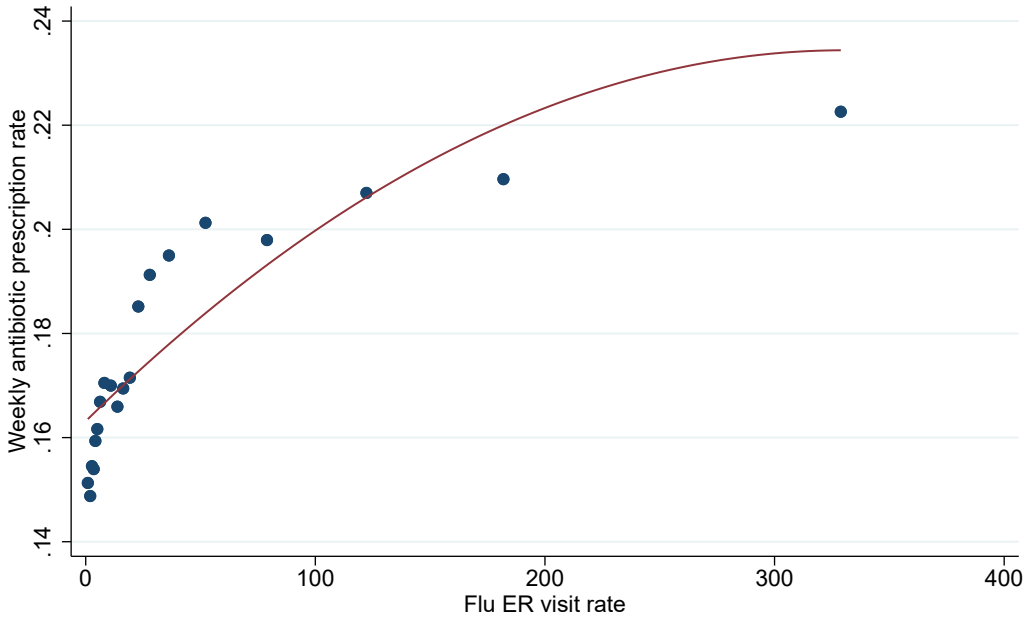


Figure 26

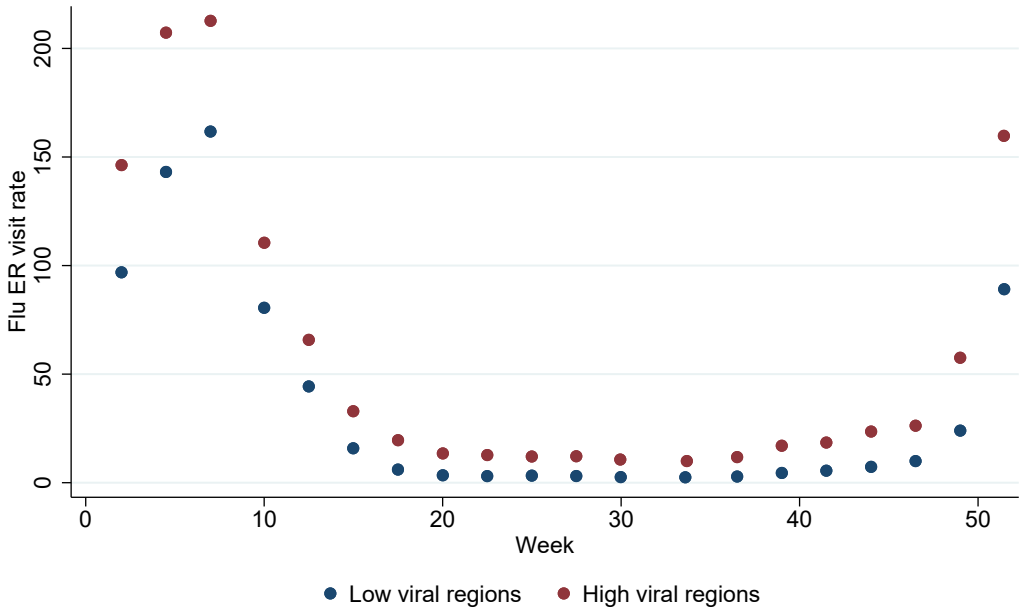


Table 13: Estimation results of Eqn. (2) with flu rates

	(1)	(2)	(3)
High Viral Region (Flu)=1	0.00309*	0.00291	-0.00244
	(0.00146)	(0.00153)	(0.0101)
Year=2015 × High Viral Region (Flu)=1	-0.00195	-0.00133	-0.0134
	(0.00176)	(0.00191)	(0.0122)
Year=2016 × High Viral Region (Flu)=1	-0.000953	-0.00147	-0.00501
	(0.00175)	(0.00190)	(0.0121)
Year=2017 × High Viral Region (Flu)=1	-0.00335*	-0.00304	-0.0179
	(0.00171)	(0.00186)	(0.0120)
Year=2018 × High Viral Region (Flu)=1	-0.00490**	-0.00482**	-0.0289*
	(0.00171)	(0.00184)	(0.0122)
Year=2019 × High Viral Region (Flu)=1	-0.00460**	-0.00418*	-0.0247*
	(0.00171)	(0.00184)	(0.0123)
Observations	80003	80002	80003

Notes: (1) Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
(2) First column reports the results from the linear probability model, second column from the weighted linear probability model and third column from the logit model.
(3)Coefficients for physician age and gender, Year, Month, Region FE's, rate of *Sometimes* class diseases and *Always* class diseases with Year FE interactions, 10th, 25th, 45th, 50th, 55th, 75th and 90th percentiles of patient age, chronic disease counts, the share of eligible patients are not reported.

Table 14: Estimation results of Eqn. (3) with continuous treatment

	LPM	Weighted LPM	Logit
Share of diseases	0.371***	0.382***	7.112***
	(0.0214)	(0.0259)	(0.326)
Year=2015 × Share of diseases	-0.00148	0.0170	-0.344
	(0.0306)	(0.0370)	(0.454)
Year=2016 × Share of diseases	-0.0637*	-0.0376	-1.140*
	(0.0289)	(0.0359)	(0.449)
Year=2017 × Share of diseases	-0.166***	-0.161***	-2.047***
	(0.0274)	(0.0331)	(0.456)
Year=2018 × Share of diseases	-0.192***	-0.167***	-2.870**
	(0.0481)	(0.0320)	(0.992)
Year=2019 × Share of diseases	-0.202***	-0.209***	-2.723***
	(0.0264)	(0.0315)	(0.445)
Observations	80003	80002	80003

Notes: (1) Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
(2) First column reports the results from the linear probability model, second column from the weighted linear probability model and third column from the logit model.
(3)Coefficients for physician age and gender, Year, Month, Region FE's, rate of *Sometimes* class diseases, 10th, 25th, 45th, 50th, 55th, 75th and 90th percentiles of patient age, chronic disease counts, the share of eligible patients are not reported.

Additional results for AAC

B Additional information on the Model

Writing the lifetime discounted utility Suppose there is only one disease group, then:

$$
\begin{aligned}
v(n,t) &= \max_{s_t^1,\ldots,s_T^1} \left\{ \sum_{t'=t}^{T} \delta^{t'-t} \cdot \mathbb{E}[U(s_{t'}^1)] + \delta^{T-t} \cdot \mathbb{E}[\beta \cdot B | n, t] \right\} \\
&= \max_{s_t^1,\ldots,s_T^1} \left\{ \sum_{t'=t}^{T} \delta^{t'-t} \cdot \mathbb{E}[U(s_{t'}^1, k')] + \delta^{T-t} \cdot \left(F(s_t^1)\, \mathbb{E}[\beta \cdot B | n, t+1] + (1 - F(s_t^1))\, \mathbb{E}[\beta \cdot B | n+1, t+1] \right) \right\} \\
&= \max_{s_t^1} \mathbb{E}[U(s_t^1)] + \delta\ F(s_t^1)\ \underbrace{\max_{s_{t+1}^1,\ldots,s_T^1} \sum_{t'=t+1}^{T} \left(\delta^{t'-(t+1)}\, \mathbb{E}[U(s_{t'}^1)] \right) + \delta^{T-(t+1)}\, \mathbb{E}[\beta \cdot B | n, t+1]}_{v(n,t+1)} \\
&\qquad + \delta\ (1 - F(s_t^1))\ \underbrace{\max_{s_{t+1}^1,\ldots,s_T^1} \sum_{t'=t+1}^{T} \left(\delta^{t'-(t+1)}\, \mathbb{E}[U(s_{t'}^1)] \right) + \delta^{T-(t+1)}\, \mathbb{E}[\beta \cdot B | n+1, t+1]}_{v(n+1,t+1)} \\
&= \max_{s_t^1} \int_0^{s_t^1} \underbrace{(b_0 + \alpha h^0(s))}_{=0} f(s) ds + \int_{s_t^1}^{1} (b + \alpha h(s)) f(s) ds + \delta \left(F(s_t^1) v(n, t+1) + (1 - F(s_t^1)) v(n+1, t+1) \right)
\end{aligned}
$$

C Additional results to Section 5.1

Table 15: Estimation results of Eqn. (7)

	OLS	IV1	IV2	OLS	IV1	IV2
E[Slope OAC]	0.000315***	0.000657***	0.000629***	0.000291***	0.000368***	0.000338***
	(0.0000773)	(0.0000941)	(0.0000924)	(0.0000782)	(0.0000994)	(0.0000982)
Year $\geq$ 2017=1 × E[Slope OAC]	0.00309***	0.00236***	0.00233***	0.00285***	0.00168***	0.00164***
	(0.000338)	(0.000337)	(0.000332)	(0.000312)	(0.000275)	(0.000271)
Passed 25% X Within 6 months around 25%=1	0.0371***	-0.0106	-0.00672	0.0394***	0.0346**	0.0384***
	(0.00151)	(0.00979)	(0.00949)	(0.00150)	(0.0118)	(0.0116)
Passed 25% X Within 6 months around 25%=1 × Year=2015	-0.00372	-0.0145*	-0.0132*	-0.00662**	-0.0179**	-0.0154**
	(0.00214)	(0.00582)	(0.00518)	(0.00217)	(0.00573)	(0.00522)
Passed 25% X Within 6 months around 25%=1 × Year=2016	0.00174	0.00361	0.00293	0.00101	0.00325	0.00220
	(0.00213)	(0.00549)	(0.00502)	(0.00215)	(0.00549)	(0.00512)
Passed 25% X Within 6 months around 25%=1 × Year=2017	0.00668**	0.0256***	0.0216***	0.00386	0.0196***	0.0172**
	(0.00211)	(0.00562)	(0.00515)	(0.00213)	(0.00578)	(0.00537)
Passed 25% X Within 6 months around 25%=1 × Year=2018	-0.000802	0.00354	0.00285	-0.000991	0.00656	0.00551
	(0.00216)	(0.00560)	(0.00520)	(0.00218)	(0.00557)	(0.00517)
Passed 25% X Within 6 months around 25%=1 × Year=2019	0.00139	0.000840	0.000981	-0.000354	0.00255	0.00309
	(0.00214)	(0.00568)	(0.00526)	(0.00214)	(0.00565)	(0.00529)
Within 6 months around 25%=1	0.0815***	0.0640***	0.0637***	0.0842***	0.0672***	0.0671***
	(0.00155)	(0.00209)	(0.00198)	(0.00159)	(0.00214)	(0.00206)
Year=2015 × Within 6 months around 25%=1	-0.00494*	0.00582	0.00509	-0.00270	0.00727*	0.00604
	(0.00218)	(0.00330)	(0.00306)	(0.00228)	(0.00335)	(0.00318)
Year=2016 × Within 6 months around 25%=1	0.00113	-0.00112	-0.000779	0.00367	0.000926	0.00146
	(0.00216)	(0.00309)	(0.00292)	(0.00221)	(0.00317)	(0.00304)
Year=2017 × Within 6 months around 25%=1	-0.00936***	-0.0159***	-0.0140***	-0.00688**	-0.0103**	-0.00909**
	(0.00210)	(0.00309)	(0.00291)	(0.00219)	(0.00327)	(0.00312)
Year=2018 × Within 6 months around 25%=1	-0.00164	-0.00232	-0.00192	-0.00158	-0.00230	-0.00172
	(0.00214)	(0.00310)	(0.00295)	(0.00220)	(0.00315)	(0.00303)
Year=2019 × Within 6 months around 25%=1	-0.00946***	-0.00562	-0.00561	-0.00871***	-0.00533	-0.00548
	(0.00214)	(0.00311)	(0.00295)	(0.00221)	(0.00321)	(0.00308)
Observations	65694	65673	65673	65694	65673	65673
idp		3.24e-135	4.67e-134		2.70e-84	6.57e-82
jp		0.0000219	0.000484		0.000000133	0.00000617

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: (1) Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
(2) IV1: Cum. incoming share of *sometimes* groups diseases interacted with year FE's and *always* group diseases without year interactions, IV2: Cum. incoming share of *always* group diseases and *sometimes* diseases interacted with year FE's.
(3) The average monthly prescription rate in the sample before 2017 was 22.2% with a standard deviation of 8.1%.
(4) The last three columns uses weights determined by the number of patients as the dependent variable is a ratio of antibiotic prescriptions to the number of patients.

C.1 Updating end-of-year performance expectations - further results

To support the hypothesis that the kink effect is due to learning as prescriptions realize across the year, I re-estimate 7 for two different groups of physicians and focusing only on the observations within 6 months interval from the event. The two group of physicians are: (i) those who are "sure" to pass 25% threshold and (ii) those who are "unsure" to pass 25% threshold. A physician is "sure" to pass 25% in a given year if for each month $m \in \{1, \cdots, 12\}$, the minimum prescription rate in his or her assigned group of (similar)

physicians (see 4.2.1) is higher than 25%.

Table 16: Estimation results of Eqn. (7)

	Sure to pass		Not sure to pass	
	OLS	IV	OLS	IV
Passed threshold=1	0.0256***	-0.0293	0.0211***	-0.0565**
	(0.00294)	(0.0257)	(0.00218)	(0.0190)
Passed threshold=1 × Year=2015	-0.00108	0.00220	0.00395	-0.00833
	(0.00438)	(0.00695)	(0.00269)	(0.00496)
Passed threshold=1 × Year=2016	0.00456	0.0000143	0.00197	-0.00735
	(0.00413)	(0.00639)	(0.00271)	(0.00497)
Passed threshold=1 × Year=2017	0.00659	0.00941	0.00952***	0.0130**
	(0.00412)	(0.00673)	(0.00268)	(0.00480)
Passed threshold=1 × Year=2018	0.117	0.142	0.00310	-0.00173
	(0.0961)	(0.0849)	(0.00259)	(0.00467)
Passed threshold=1 × Year=2019	0.0107	-0.00126	0.00500	0.00761
	(0.114)	(0.127)	(0.00263)	(0.00468)
Observations	5724	5724	23283	23283

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: (1) Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
(2) IV1: Cum. incoming share of *sometimes* groups diseases interacted with year FE's and *always* group diseases without year interactions, IV2: Cum. incoming share of *always* group diseases and *sometimes* diseases interacted with year FE's.